

ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO



FACULTAD DE CIENCIAS

ESCUELA DE FÍSICA Y MATEMÁTICA

INGENIERÍA EN ESTADÍSTICA INFORMÁTICA

**“MODELACIÓN Y PRONÓSTICO DEL NÚMERO DE ABONADOS Y EL
CONSUMO DE ENERGÍA ELÉCTRICA EN MWH DE LA EMPRESA ELÉCTRICA
RIOBAMBA S.A.”**

**TESIS DE GRADO PREVIA A LA OBTENCIÓN DEL TÍTULO DE
INGENIERA EN ESTADÍSTICA INFORMÁTICA**

PRESENTADO POR:

VALERIA NATALY VACACELA COLCHA

Riobamba – Ecuador

2014

DEDICATORIA.

A Dios por guiarme por el buen camino con valentía.

A mis padres Fausto y Magdalena, a mis hermanos Diego y Lorena por su apoyo incondicional, consejos, comprensión y amor.

A mis hijos David y Mauricio quienes siempre serán mi motivación, inspiración y felicidad.

Valeria Nataly Vacacela Colcha.

AGRADECIMIENTO.

Agradezco a Dios por bendecirme y hacer realidad este sueño anhelado.

A la ESPOCH por darme la oportunidad de ser una profesional.

A mi tutor de tesis, Dr. Jorge Congacha y colaboradora Ing. Isabel Escudero, quienes con sus conocimientos, experiencia, y paciencia motivaron la culminación de ésta etapa de mi vida.

A mis profesores, familiares y amigos quienes aportaron con un granito de arena a mi formación con sus consejos, enseñanzas, compañía y sobre todo su por su amistad sincera.

ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO

FACULTAD DE CIENCIAS

ESCUELA DE FÍSICA Y MATEMÁTICA

El Tribunal de Tesis certifica que: El trabajo de investigación: **“MODELACIÓN Y PRONÓSTICO DEL NÚMERO DE ABONADOS Y EL CONSUMO DE ENERGÍA ELÉCTRICA EN MWH DE LA EMPRESA ELÉCTRICA RIOBAMBA S.A.”**, de responsabilidad de la señorita egresada Valeria Nataly Vacacela Colcha, ha sido prolijamente revisado por los Miembros del Tribunal de Tesis, quedando autorizada su presentación.

FIRMA

FECHA

Ing. Cesar Avalos.

DECANO FAC. CIENCIAS

Dra. Jenny Orbe

DIRECTORA DE ESCUELA

Dr. Jorge Congacha

DIRECTOR DE TESIS

Ing. Isabel Escudero

MIEMBRO DE TRIBUNAL

DIRECTOR CENTRO

DE DOCUMENTACIÓN

NOTA DE TESIS ESCRITA_____

Yo **Valeria Nataly Vacacela Colcha**, soy responsable de las ideas, doctrinas y resultados expuestos en esta Tesis; y el patrimonio intelectual de la Tesis de Grado, pertenece a la ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO.

VALERIA NATALY VACACELA COLCHA.

ÍNDICE DE ABREVIATURAS

EERSA	Empresa Eléctrica Riobamba Sociedad Anónima.
MWH	Megawatt-hora
MVA	Megavoltiamperio
kW	Kilowatt
kV	Kilovoltio
V	Voltio
AED	Análisis Exploratorio de Datos
Km	Kilómetros
MAD	Mean Absolute Desviation
MSE	Mean Squared Error
MAPE	Mean Absolute Porcentage Error
H_0	Hipótesis Nula
H_1	Hipótesis Alternativa
Q_1	Primer Cuartil
Q_3	Tercer Cuartil
IQR	Rango Intercuartil
NACGNVB	Número de Abonados Categoría General Nivel de Voltaje Bajo
NACGNVM	Número de Abonados Categoría General Nivel de Voltaje Medio
NACR	Número de Abonados Categoría Residencial
CECAP	Consumo de Energía en MWh Categoría Alumbrado Público.
CECGNVB	Consumo de Energía en MWh Categoría General Nivel Voltaje Bajo
CECGNVM	Consumo de Energía en MWh Categoría General Nivel Voltaje Medio
CECGNVA	Consumo de Energía en MWh Categoría General Nivel Voltaje Alto
CECR	Consumo de Energía en MWh Categoría Residencial

ÍNDICE GENERAL

ÍNDICE DE ABREVIATURAS.....	i
ÍNDICE DE TABLAS.	vi
ÍNDICE DE GRÁFICOS	vii
ÍNDICE DE ILUSTRACIONES	x
ÍNDICE DE ANEXOS.....	xi
INTRODUCCIÓN	14
OBJETIVOS	15
OBJETIVO GENERAL	15
OBJETIVOS ESPECÍFICOS	15
CAPÍTULO I: FUNDAMENTACIÓN TEÓRICA	16
1.1 LA ENERGÍA ELÉCTRICA EN EL ECUADOR	16
1.1.1 DISTRIBUCIÓN DE ENERGÍA.....	17
1.1.2 SISTEMA DE SUBTRANSMISIÓN	18
1.1.3 CATEGORÍAS ESTABLECIDAS A PARTIR DEL AÑO 2009	18
1.1.3.1 Categoría Residencial	18
1.1.3.2 Categoría General.....	19
1.1.3.3 Categoría Alumbrado Público:.....	19
1.1.4 NIVELES DE TENSIÓN.	20
1.1.4.1 Alta Tensión.....	20
1.1.4.2 Media Tensión.....	20
1.1.4.3 Baja Tensión	20
1.2 ANÁLISIS EXPLORATORIO DE DATOS (AED)	20
1.2.1 DATOS ATÍPICOS.....	21
1.2.1.1 Causas de la Presencia de Datos Atípicos	21
1.2.1.2 Procedimientos para identificar datos atípicos.....	21
1.2.2 ESTUDIO DE LA NORMALIDAD	23

1.3	PRONÓSTICOS	24
1.3.1	MÉTODOS DE PRONÓSTICOS.....	24
1.4	SERIES TEMPORALES.....	26
1.4.1	COMPONENTES DE SERIES TEMPORALES	26
1.5	LAS TÉCNICAS MÁS UTILIZADAS EN PRONÓSTICOS DE SERIES DE TIEMPO	27
1.5.1	SUAVIZACIÓN EXPONENCIAL.....	27
1.5.1.1	Suavización Exponencial Simple (Método de Brown)	27
1.5.1.2	Suavización Exponencial Doble (Método de Holt).....	27
1.5.1.3	Suavización Exponencial Triple (Método de Winters)	28
1.5.2	PROMEDIO MÓVIL	30
1.6	ERROR DE PRONÓSTICO O RESIDUAL	30
1.7	MEDIDAS DE PRECISIÓN	30
1.7.1	MEAN ABSOLUTE DESVIATION (MAD).....	31
1.7.2	MEAN SQUARED ERROR (MSE)	31
1.7.3	MEAN ABSOLUTE PORCENTAGE ERROR (MAPE).....	32
1.8	PRUEBA DE HIPÓTESIS	32
1.8.1	ERRORES DE TIPO I Y TIPO II.....	33
1.8.2	PASOS PARA REALIZAR UNA PRUEBA DE HIPÓTESIS.....	34
	CAPÍTULO II: METODOLOGÍA DE LA INVESTIGACIÓN	36
2.1	LUGAR DE LA INVESTIGACIÓN	36
2.2	MÉTODOS Y TÉCNICAS.....	36
2.2.1	ETAPAS DEL AED	36
2.2.1.1	Preparación de Datos	36
2.2.1.2	Identificación de Datos Atípicos	37
2.2.1.3	Prueba de Normalidad	38
2.2.2	ELECCIÓN DE UNA TÉCNICA DE PRONÓSTICO.....	39
2.2.3	PRUEBA DE HIPÓTESIS	40
	CAPÍTULO III: ANÁLISIS DE DATOS Y RESULTADOS	42

3.1	IDENTIFICACIÓN DE POSIBLES DATOS ATÍPICOS EN EL PERIODO 2009-2012	42
3.1.1	MÉTODO GRÁFICO.....	42
3.1.1.1	BOX-PLOT del Número de Abonados	42
3.1.1.2	BOX-PLOT del Consumo de Energía en MWh	44
3.1.2	MÉTODO ANALÍTICO	45
3.2	DEPURACIÓN DE DATOS Y PRUEBA DE NORMALIDAD	45
3.2.1	PRUEBA DE KOLMOGOROV SMIRNOV PARA EL NUMERO DE ABONADOS	46
3.2.2	PRUEBA DE KOLMOGOROV SMIRNOV PARA EL CONSUMO DE ENERGÍA EN MWh.....	47
3.3	TRANSFORMACIONES DE LA VARIABLE QUE NO SE AJUSTA A LA NORMAL	50
3.4	EXPLORACIÓN DE LAS COMPONENTES DE LAS SERIES TEMPORALES.....	51
3.4.1	NÚMERO DE ABONADOS EN EL PERIODO 2009-2012.....	51
3.4.2	GRÁFICAS DE SERIES DE TIEMPO Y AUTOCORRELOGRAMAS PARA DATOS DEL CONSUMO DE ENERGÍA EN MWh EN EL PERIODO 2009-2012	53
3.5	IDENTIFICACIÓN DEL MODELO DE PRONÓSTICO PARA DATOS DEL PERIODO 2009-2012	56
3.5.1	MODELOS PARA DATOS DEL NÚMERO DE ABONADOS.....	56
3.5.2	MODELOS PARA DATOS DEL CONSUMO DE ENERGÍA EN MWh.....	58
3.6	ESTIMACIÓN DE LOS PARÁMETROS DEL LOS MODELOS.....	60
3.7	VALUACIÓN DEL MODELO	63
3.7.1	EVALUACIÓN DEL MODELO PARA DATOS DEL NÚMERO DE ABONADOS	63
3.7.2	EVALUACIÓN DEL MODELO PARA DATOS DEL CONSUMO DE ENERGÍA EN MWH	64

3.8	UTILIZACIÓN DEL MODELO PARA GENERAR PRONÓSTICOS.....	65
3.8.1	UTILIZACIÓN DEL MODELO PARA DATOS DEL NÚMERO DE ABONADOS	65
3.8.2	UTILIZACIÓN DEL MODELO PARA DATOS DEL CONSUMO DE ENERGÍA EN MWh.....	66
3.9	VALIDACIÓN DEL MODELO	68
3.9.1	VALIDACIÓN DEL MODELO PARA DATOS DEL NÚMERO DE ABONADOS.	68
3.9.2	VALIDACIÓN DEL MODELO PARA DATOS DEL CONSUMO DE ENERGÍA EN MWh.....	69
3.10	PRUEBA DE HIPÓTESIS	70
3.10.1	PRUEBA DE HIPÓTESIS PARA EL NÚMERO DE ABONADOS DE LA CATEGORÍA GENERAL NIVEL DE VOLTAJE MEDIO.....	70
3.10.2	PRUEBA DE HIPÓTESIS PARA EL NÚMERO DE ABONADOS DE LA CATEGORÍA RESIDENCIAL.	72
3.10.3	PRUEBA DE HIPÓTESIS PARA CONSUMO DE ENERGÍA EN MWh DE LA CATEGORÍA GENERAL NIVEL DE VOLTAJE MEDIO.....	74
3.10.4	PLANTEAMIENTO DE HIPÓTESIS PARA CONSUMO DE ENERGÍA EN MWh DE LA CATEGORÍA GENERAL NIVEL DE VOLTAJE ALTO.....	76
	CAPÍTULO IV: CONCLUSIONES Y RECOMENDACIONES	79
4.1	CONCLUSIONES.	79
4.2	RECOMENDACIONES.	81
	RESUMEN	82
	SUMMARY	83
	BIBLIOGRAFÍA	84
	ANEXOS	86

ÍNDICE DE TABLAS.

Tabla 1. SELECCIÓN DEL ESTADÍSTICO DE PRUEBA.	34
Tabla 2. IDENTIFICACIÓN DE DATOS ATÍPICOS LEVES	38
Tabla 3. IDENTIFICACIÓN DE DATOS ATÍPICOS EXTREMOS	38
Tabla 4. MODELOS PARA EL NACGNVB	56
Tabla 5. MODELOS PARA EL NACGNVM.....	57
Tabla 6. MODELOS PARA EL NACR.....	57
Tabla 7. MODELOS PARA EL CECAP	58
Tabla 8. MODELOS PARA EL CECGNVB	58
Tabla 9. MODELOS PARA EL CECGNVM	59
Tabla 10. MODELOS PARA EL CECGNVA.....	59
Tabla 11. MODELOS PARA EL CECR	60
Tabla 12. PRONÓSTICOS DEL NÚMERO DE ABONADO PARA EL AÑO 2013 ..	65
Tabla 13. MEDIDAS DESCRIPTIVAS DE LOS PRONÓSTICOS DEL NÚMERO DE ABONADOS PARA EL AÑO 2013	66
Tabla 14. PRONOSTICO DEL CONSUMO DE ENERGÍA EN MWH PARA EL AÑO 2013	66
Tabla 15. MEDIDAS DESCRIPTIVAS DE LOS PRONÓSTICOS DEL CONSUMO DE ENERGÍA EN MWh PARA EL AÑO 2013.	67
Tabla 16. NACGNVM REALES Y PRONÓSTICOS.....	71
Tabla 17. PRUEBA T PARA EL NACGNVM.....	71
Tabla 18. NABCR REALES Y PRONÓSTICOS	73
Tabla 19. PRUEBA T PARA EL NACR	73
Tabla 20. CECGNVM REALES Y PRONÓSTICOS	75
Tabla 21. PRUEBA T PARA EL CECGNVM REALES Y PRONÓSTICOS	75
Tabla 22. CECGNVA REALES Y PRONÓSTICOS.....	77
Tabla 23. PRUEBA T PARA EL CECGNVA REALES Y PRONÓSTICOS	77

ÍNDICE DE GRÁFICOS

Gráfica 1. NÚMERO DE ABONADOS CATEGORÍA GENERAL EN EL NIVEL DE VOLTAJE BAJO Y MEDIO.	42
Gráfica 2. NÚMERO DE ABONADOS CATEGORÍA RESIDENCIAL.....	43
Gráfica 3. CONSUMO DE ENERGÍA EN MWh EN LA CATEGORÍA ALUMBRADO PÚBLICO.	44
Gráfica 4. CONSUMO DE ENERGIA EN MWh DE LA CATEGORIA GENERAL	44
Gráfica 5. CONSUMO DE ENERGÍA EN MWh DE LA CATEGORÍA RESIDENCIAL	45
Gráfica 6. PRUEBA DE NORMALIDAD PARA EL NACGNVB	46
Gráfica 7. PRUEBA DE NORMALIDAD PARA EL NACGNVM	46
Gráfica 8. PRUEBA DE NORMALIDAD PARA EL NACR	47
Gráfica 9. PRUEBA DE NORMALIDAD PARA EL CECAP	47
Gráfica 10. PRUEBA DE NORMALIDAD PARA EL CECGNVB	48
Gráfica 11. PRUEBA DE NORMALIDAD PARA EL CECGNVM	48
Gráfica 12. PRUEBA DE NORMALIDAD PARA EL CECGNVA	49
Gráfica 13. PRUEBA DE NORMALIDAD PARA EL CECR.....	49
Gráfica 14. NORMALIDAD DE DATOS TRANSFORMADOS DEL CECGNV	50
Gráfica 15. NACGNVB	51
Gráfica 16. FUNCIÓN DE AUTOCORRELACIÓN PARA EL NACGNVB	51
Gráfica 17. NACGNVM	52
Gráfica 18. FUNCIÓN DE AUTOCORRELACIÓN PARA EL NACGNVM	52
Gráfica 19. NACR	52
Gráfica 20. FUNCIÓN DE AUTOCORRELACIÓN PARA EL NACR	52
Gráfica 21. SERIES DE TIEMPO CECAP.....	53
Gráfica 22. FUNCIÓN DE AUTOCORRELACIÓN PARA EL CECAP.....	53
Gráfica 23. SERIES DE TIEMPO CECGNVB.....	53

Gráfica 24. FUNCIÓN DE AUTOCORRELACIÓN PARA EL CECGNVB	53
Gráfica 25. SERIES DE TIEMPO CECGNVM.....	54
Gráfica 26. FUNCIÓN DE AUTOCORRELACIÓN PARA EL CECGNVM.....	54
Gráfica 27. SERIES DE TIEMPO CECGNVA	54
Gráfica 28. FUNCIÓN DE AUTOCORRELACIÓN PARA EL CECGNVA	54
Gráfica 29. FUNCIÓN DE AUTOCORRELACIÓN DE LAS DIFERENCIAS PARA EL CECGNVA.....	55
Gráfica 30. SERIES DE TIEMPO CECR.....	56
Gráfica 31. FUNCIÓN DE AUTOCORRELACIÓN PARA EL CECR	56
Gráfica 32. FUNCIÓN DE AUTOCORRELACIÓN PARA LOS RESIDUOS DEL NACGNVB	63
Gráfica 33. FUNCIÓN DE AUTOCORRELACIÓN PARA LOS RESIDUOS DEL NACGNVM	63
Gráfica 34. FUNCIÓN DE AUTOCORRELACIÓN PARA LOS RESIDUOS DEL NACR	63
Gráfica 35. FUNCIÓN DE AUTOCORRELACIÓN PARA LOS RESIDUO DEL CECAP.....	64
Gráfica 36. FUNCIÓN DE AUTOCORRELACIÓN PARA LOS RESIDUO DEL CECGNVB	64
Gráfica 37. FUNCIÓN DE AUTOCORRELACIÓN PARA LOS RESIDUO DEL CECGNVM.....	64
Gráfica 38. FUNCIÓN DE AUTOCORRELACIÓN PARA LOS RESIDUO DEL CECGNVA.....	65
Gráfica 39. FUNCIÓN DE AUTOCORRELACIÓN PARA LOS RESIDUOS DEL CECR	65
Gráfica 40. RESIDUOS DEL NACGNVB	68
Gráfica 41. RESIDUOS DEL NACGNVM	68
Gráfica 42: RESIDUOS DEL NACR	68
Gráfica 43: RESIDUOS VARIABLE CECAP	69

Gráfica 44: RESIDUOS DEL CECGNVB	69
Gráfica 45: RESIDUOS DEL CECGNVM.....	69

ÍNDICE DE ILUSTRACIONES

Ilustración 1: SUMINISTRO Y MONTAJE DE TRES TRANSFORMADORES DE POTENCIA EN LAS SUBESTACIONES 1, 4, 7.....	16
Ilustración 2: PARTES DEL BOX-PLOT	22
Ilustración 3: MÉTODOS DE PRONÓSTICOS.....	25
Ilustración 4: COMPONENTES DE SERIE TEMPORALES.....	26
Ilustración 5: ZONA DE ACEPTACIÓN Y RECHAZO DE LA DISTRIBUCIÓN NORMAL $\alpha=0.05$	35
Ilustración 6: ORGANIZACIÓN DE LOS DATOS.	37
Ilustración 7: MÉTODO GRÁFICO.....	38
Ilustración 8: PRUEBA DE NORMALIDAD	39
Ilustración 9: REGIONES DE ACEPTACIÓN Y RECHAZO DE LA PRUEBA T- STUDENT	41

ÍNDICE DE ANEXOS

Anexo 1	MATRIZ DE DATOS DEL NÚMERO DE ABONADOS DEL PERIODO 2009-2012.....	86
Anexo 2	MATRIZ DE DATOS DEL CONSUMO DE ENERGÍA EN MWH DEL PERIODO 2009-2012	87
Anexo 3	MATRIZ DE DATOS DEL NÚMERO DE ABONADOS DEL AÑO 2013	88
Anexo 4	MATRIZ DE DATOS DEL CONSUMO DE ENERGÍA DEL AÑO 2013	89

INTRODUCCIÓN

La parte eléctrica ha sufrido tres significativos cambios, un primer cambio se da con la creación del Instituto Ecuatoriano de Electrificación INECEL , con el objetivo de planificar, ejecutar, operar, regular y controlar la actividad del sector eléctrico; el segundo cambio se da con la elaboración del Primer Plan Nacional de Electrificación y se publica la Ley de Régimen del Sector Eléctrico, con el objetivo de satisfacer las necesidades de energía eléctrica del país, haciendo uso correcto y aprovechamiento de los recursos naturales; y el tercer cambio que comienza con el presidente Rafael Correa, tomando como responsable del manejo y gestión de las áreas estratégicas.

La energía eléctrica ha sido un elemento evolucionario de la industria, la medicina, la ciencia, el comercio, la tecnología, etc., convirtiéndose así en uno de los servicios de mayor importancia para el desarrollo de las ciudades y de la humanidad. Desde 1963 inicia la comercialización de la energía eléctrica en la provincia de Chimborazo, teniendo como política brindar el mejor servicio básico a la provincia y tiene una cobertura del 97%. La EERSA cuenta con tres Centrales Hidroeléctricas que generan energía propia; Central Alao, Central Río Blanco y la de Nizag.

Hoy en día la energía eléctrica es una necesidad de la cual no podemos prescindir y que va en aumento. Por este motivo hay que conseguir generar modelos de pronósticos que permitirá conocer el consumo y el número de abonados a futuro. Por medio del AED y componentes de series temporales se establecen modelos para cada variable.

OBJETIVOS

OBJETIVO GENERAL

Modelar y predecir el número de abonados y el consumo de energía eléctrica en función de categorías y nivel de voltaje.

OBJETIVOS ESPECÍFICOS

- Realizar un análisis exploratorio de datos (AED) mensuales del número de abonados y el consumo de energía por categoría y nivel de voltaje.
- Depurar la información extraída del AED.
- Identificar el modelo de predicción.
- Estimar los parámetros del modelo.
- Predecir el número de abonados, consumo de energía por categoría y nivel de voltaje.
- Evaluar el modelo.

CAPÍTULO I

FUNDAMENTACIÓN TEÓRICA

1.1 LA ENERGÍA ELÉCTRICA EN EL ECUADOR



Ilustración 1: SUMINISTRO Y MONTAJE DE TRES TRANSFORMADORES DE POTENCIA EN LAS SUBESTACIONES 1, 4, 7

Una de las particularidades de nuestro querido Ecuador es la irregular topografía, con varios tipos de climas y ambientes ideales que le otorgan un alto potencial de energías renovables y limpias, las condiciones de cobertura y satisfacción de la demanda que se presentan en la actualidad, demuestran un estrecho vínculo especialmente con la electrificación y energización rural.

El desarrollo del sector energético es estratégico para el Ecuador. En los últimos 15 años el país ha experimentado una fuerte dependencia de combustibles fósiles generando un cambio

notable en su matriz energética, la energía eléctrica producida de fuentes térmicas equivale al 43,3% en la actualidad, mientras las provenientes de fuentes hidroeléctricas es de 45,3%.

El Ecuador ha presentado en los últimos 20 años un desequilibrio entre oferta y demanda de energía eléctrica. Por un lado el aumento de la demanda ha incrementado alrededor de un 6,5% anual, mientras que la oferta de energía únicamente se ha incrementado en un 5%. El Ecuador actualmente produce energía eléctrica por medio de centrales hidroeléctricas y termoeléctricas. La empresa pública encargada de la generación y transmisión de la energía eléctrica en el Ecuador es la CELEC (Corporación Eléctrica del Ecuador).(5)

1.1.1 DISTRIBUCIÓN DE ENERGÍA

A partir de Enero del 2009, tras la fusión de 10 regionales para conformar la Corporación Nacional de Electricidad (CNEL), las 11 empresas de distribución de energía eléctrica del Ecuador son:

- Empresa Eléctrica Ambato Regional Centro Norte S.A.
- Empresa Eléctrica Azogues C.A.
- Empresa Eléctrica Regional Centro Sur C.A.
- Empresa Eléctrica Provincial Cotopaxi S.A.
- Empresa Eléctrica Provincial Galápagos S.A.
- Empresa Eléctrica Regional Norte S.A.
- Empresa Eléctrica Quito S.A.
- Empresa Eléctrica Riobamba S.A.
- Empresa Eléctrica Regional de Sur S.A.
- Unidad Eléctrica de Guayaquil.
- Corporación Nacional de Electricidad (CNEL) S.A.(6)

1.1.2 SISTEMA DE SUBTRANSMISIÓN

El sistema eléctrico de la Empresa Eléctrica Riobamba, se encuentra formado por el sistema de subtransmisión a 69 kV el cual está conectado al Sistema Nacional Interconectado, a través de la Subestación Riobamba perteneciente a TRANSELECTRIC.

En lo referente al sistema de distribución, a través de las 10 Subestaciones de distribución se extienden 31 alimentadores primarios, con una longitud total de 2.644 km, de los cuales el 79% tiene configuración monofásica, el 18% es configuración trifásica y el 3% es bifásica.

En redes secundarias, se cuenta con una longitud total de 4.720 km de redes de baja tensión, de las cuales el 80% son circuitos monofásicos, el 19% bifásicos y el 1% son circuitos trifásicos. En relación al alumbrado público se cuenta con un total de 20.681 luminarias, con una potencia total instalada de 3.183 kW

1.1.3 CATEGORÍAS ESTABLECIDAS A PARTIR DEL AÑO 2009

1.1.3.1 Categoría Residencial

Corresponde al servicio eléctrico destinado exclusivamente al uso doméstico de los consumidores, es decir, en la residencia de la unidad familiar independientemente de la cantidad de electrodomésticos conectados. También se incluye a los consumidores de escasos recursos económicos y bajos consumos que tienen integrada a su vivienda una pequeña actividad comercial o artesanal.

1.1.3.2 Categoría General

Básicamente comprenden el comercio, la industria y la prestación de servicios públicos y privados.

Se consideran dentro de esta categoría, entre otros, los siguientes:

- a) Locales y establecimientos comerciales públicos o privados.
- b) Locales públicos o privados destinados a la elaboración o transformación de productos por medio de cualquier proceso industrial y sus oficinas administrativas.
- c) Unidades Educativas.
- d) Instalaciones de bombeo de agua.
- e) Entidades de asistencia social.
- f) Entidades de beneficio público.
- g) Entidades oficiales.
- h) Escenarios deportivos.
- i) Cultos religiosos.
- j) Servicio comunitario.

1.1.3.3 Categoría Alumbrado Público:

Se aplica a los consumos destinados al alumbrado de calles, avenidas y en general, vías de circulación pública, a la iluminación de plazas, parques, monumentos de propiedad pública, sistemas ornamentales de fuentes públicas y a los sistemas de señalamiento luminoso utilizados para el control del tránsito.

1.1.4 NIVELES DE TENSIÓN.

1.1.4.1 Alta Tensión

Para voltajes de suministro en el punto de entrega mayores a 40000V y asociados con la Subtransmisión.

1.1.4.2 Media Tensión

Para voltajes de suministro en el punto de entrega entre 600V y 40000V. Dentro de este grupo se incluyen los consumidores que se conectan a la red de media tensión a través de Transformadores de Distribución de propiedad de la EERSA para uso exclusivo o de propiedad del consumidor.

1.1.4.3 Baja Tensión

Para voltajes de suministro en el punto de entrega inferiores a 600V.(7)

1.2 ANÁLISIS EXPLORATORIO DE DATOS (AED)

El AED es considerado actualmente como un conjunto de procedimientos cuyo objetivo es proporcionar una visión detallada de las variables cuyo nivel de medición es el intervalo. Se apoya en un planteamiento descriptivo y se realiza sin aceptar ideas preconcebidas sobre el contenido informativo de los datos, o lo que es lo mismo, con una mentalidad “exploratoria”.

Desde un punto de vista técnico, el AED se caracteriza por el empleo de procedimientos analíticos y descriptivos de carácter gráfico o semigráfico ya que éstos muestran particularidades y caracteres de las variables mostrando las estructuras ocultas de los datos.

1.2.1 DATOS ATÍPICOS

Un dato atípico en un conjunto de datos, es una observación que es lejana, en valor, del resto de datos; es decir, es un dato inusualmente grande o inusualmente pequeño, comparado con los demás. Un valor atípico puede ser el resultado de un error en una medición, en cuyo caso distorsiona la interpretación de los datos al tener influencia excesiva sobre los cálculos a partir de la muestra. Por esta razón, todos los valores atípicos deben ser examinados cuidadosamente antes de realizar un análisis formal y no se los debería eliminar sin una justificación previa.

1.2.1.1 Causas de la Presencia de Datos Atípicos

Los datos atípicos pueden presentarse en las siguientes causas.

La primera causa es cuando se genera un error en el procedimiento, es decir al ingresar los datos o en la codificación.

La segunda causa es al mirar de forma subjetiva los resultados, en este caso se puede eliminar del análisis.

La tercera causa es en la medición de los datos que están dentro del rango pero tienen son únicas en las variables.

La cuarta causa surge de datos extraordinarios que no tienen explicación, sin embargo se debe reportarlo y generar conclusiones. (8) (9)

1.2.1.2 Procedimientos para identificar datos atípicos

Los procedimientos para la identificación de datos atípicos es el método gráfico Box-plot y el método de cuartiles.

Box-plot.

Proporcionan una descripción general del comportamiento de los datos, identificando así datos atípicos.

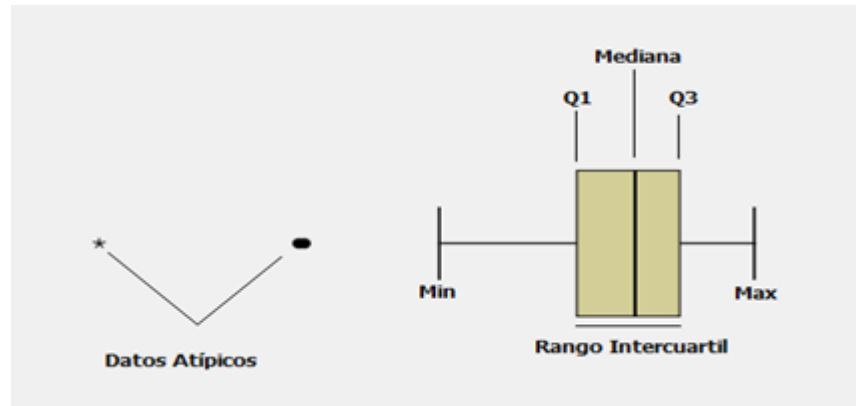


Ilustración 2: PARTES DEL BOX-PLOT

Método de Cuartiles.

Debido a que los cuartiles dividen a la distribución en 4 partes por lo tanto se vuelve una medida insensible ante la presencia de datos atípicos.

Valor atípico leve.

Se considera un valor leve aquel que se encuentra distante en relación a los demás datos, cumple las siguientes condiciones:

$$Q_1 - 1.5 IQR < Y_t < Q_3 + 1.5 IQR \quad (1)$$

Donde:

Y_t = observación en el periodo t.

Q_1 = primer cuartil.

Q_3 = tercer cuartil.

IQR = rango intercuartil ($Q_3 - Q_1$).

Valor atípico extremo.

Un valor atípico extremo es aquel que sale de los límites:

$$Q_1 - 3 IQR < Y_t < Q_3 + 3 IQR \quad (2)$$

Donde:

Y_t = observación en el periodo t.

Q_1 = primer cuartil.

Q_3 = tercer cuartil.

IQR = rango intercuartil ($Q_3 - Q_1$).

1.2.2 ESTUDIO DE LA NORMALIDAD

Para realizar análisis estadísticos es indispensable verificar la normalidad de los datos. Una forma para comprobar la normalidad de una distribución es efectuar la prueba de Shapiro Wilks si las muestras son pequeñas ($n < 30$), y la prueba de Kolmogorov-Smirnov si las muestras son grandes.

En caso de que los datos no sigan una distribución normal se puede considerar varias transformaciones como:

- $y^* = \ln(y)$
- $y^* = \log(y)$
- $y^* = 1/y$
- $y^* = \sqrt{y}$
- $y^* = y^2$

- $y^* = 1/\sqrt{y} \quad (4)(10)$

En donde:

y^* = observación transformada.

y = observación o valor real.

\ln = logaritmo natural.

\log = logaritmo en base 10.

1.3 PRONÓSTICOS

Es el proceso en situaciones de incertidumbre, van asociados a diferentes usos y características propias de las variables en estudio, por ello se debe ser cuidadoso con el método de pronóstico seleccionado ya que dependen del ajuste de los errores.

Las características a considerar son:

- Mejorar los pronósticos para reducir dicho error.
- Aumentar la flexibilidad de las operaciones.
- Reducir el tiempo de anticipación de los pronósticos.

Todo pronóstico tiene un margen de error y el objetivo es minimizarlo.

1.3.1 MÉTODOS DE PRONÓSTICOS

Los métodos de pronósticos son métodos cualitativos y cuantitativos:

Métodos Cualitativos, esto se basa fundamentalmente en el uso de la experiencia, intuición y buen juicio de los que elaboran los pronósticos.

Métodos Cuantitativos, hacen uso de la estadística para analizar la información histórica, para la elaboración de las proyecciones.

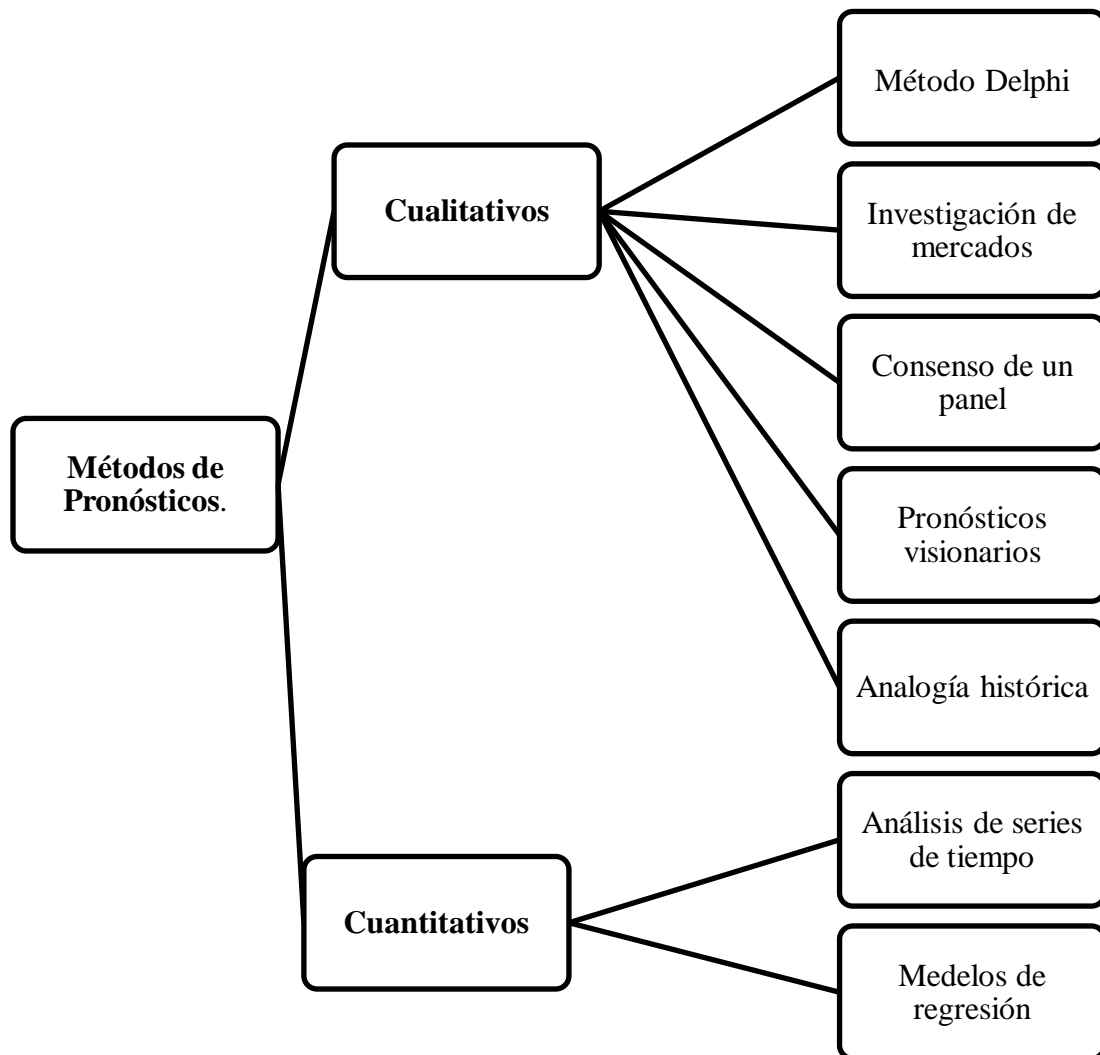


Ilustración 3: MÉTODOS DE PRONÓSTICOS

1.4 SERIES TEMPORALES

Es una secuencia de datos medidos en determinados momentos y ordenados cronológicamente.

1.4.1 COMPONENTES DE SERIES TEMPORALES

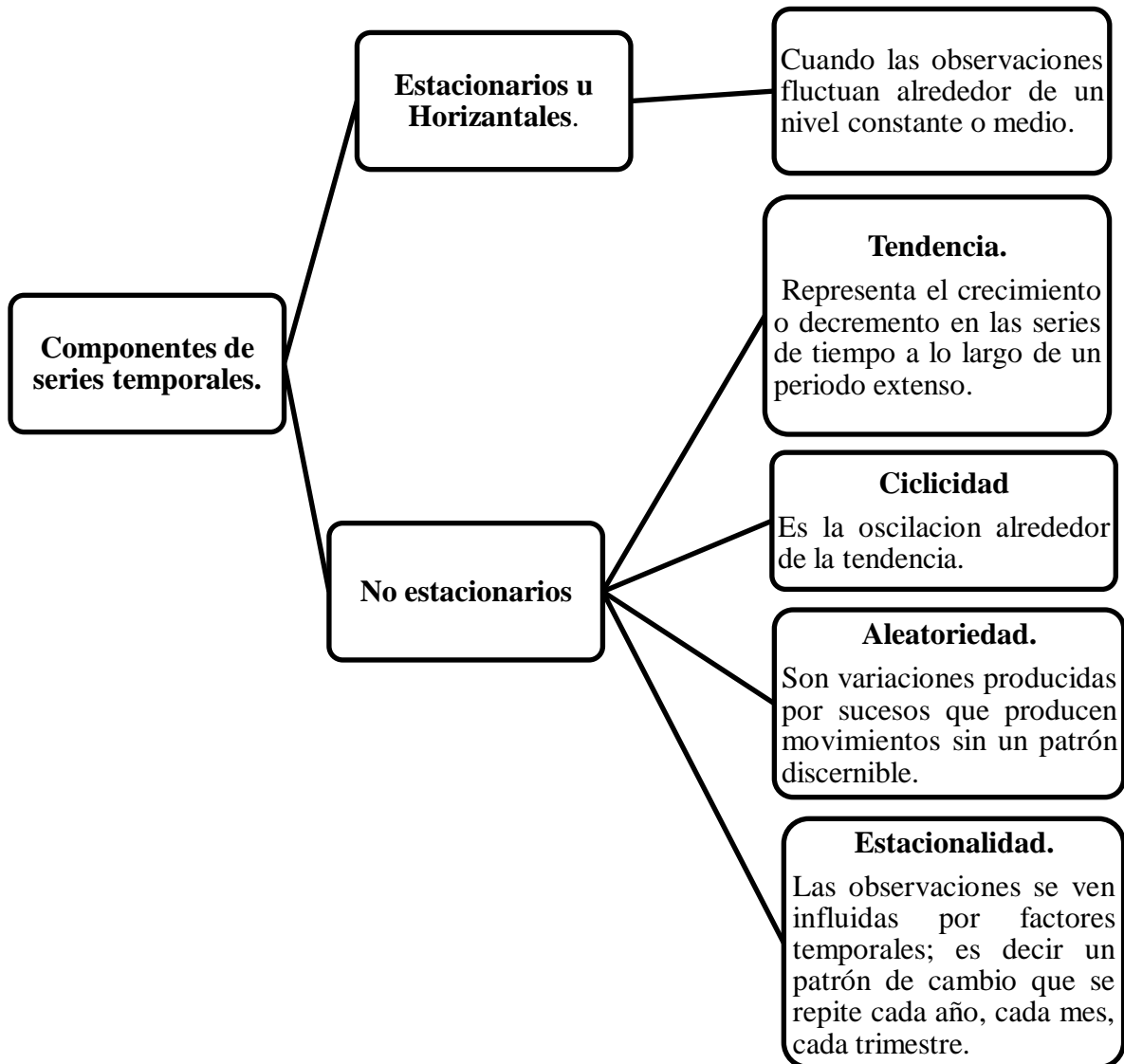


Ilustración 4: COMPONENTES DE SERIE TEMPORALES.

1.5 TÉCNICAS MÁS UTILIZADAS EN PRONÓSTICOS DE SERIES DE TIEMPO

1.5.1 SUAVIZACIÓN EXPONENCIAL

1.5.1.1 Suavización Exponencial Simple (Método de Brown)

El suavizamiento exponencial simple proporciona un promedio móvil con un peso exponencial de todos los valores observados con anterioridad, con frecuencia, el modelo es apropiado para datos con tendencia no predecible; la ecuación de suavizamiento exponencial es:

$$\hat{Y}_{t+1} = \alpha Y_t + (1 - \alpha) \hat{Y}_t \quad (3)$$

Donde

\hat{Y}_{t+1} = nuevo valor suavizado o el valor del pronóstico para el siguiente periodo.

α = constante de suavizamiento ($0 < \alpha < 1$).

Y_t = nueva observación o valor real de una serie en el periodo t.

\hat{Y}_t = antigua valor suavizado o pronostico para el periodo t.

1.5.1.2 Suavización Exponencial Doble (Método de Holt)

En algunas situaciones los datos observados tienen tendencia clara, información que permita anticipar movimientos futuros estimando así el nivel actual y la pendiente, cuando esto sucede es necesario una función de pronóstico de tendencia lineal.

Las tres ecuaciones que se usan en el método de Holt son:

1. La serie suavizada exponencialmente, o estimado del nivel actual:

$$L_t = \alpha Y_t + (1 - \alpha)(L_{t-1} + T_{t-1}) \quad (4)$$

2. El estimado de la tendencia:

$$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1} \quad (5)$$

3. Pronóstico del periodo p en el futuro:

$$\hat{Y}_{t+p} = L_t + pT_t \quad (6)$$

Donde:

L_t = nuevo valor suavizado (estimado del nivel actual).

α = constante de suavizamiento para el nivel ($0 < \alpha < 1$).

Y_t = Observación nueva o valor real de la serie en el periodo t .

β = constante de suavizamiento para el estimado de tendencia ($0 < \beta < 1$).

T_t = estimado de tendencia.

p = periodos a pronosticarse en el futuro.

\hat{Y}_{t+p} = pronósticos para el periodo p en el futuro.

1.5.1.3 Suavización Exponencial Triple (Método de Winters)

Incorpora influencias estacionales en el pronóstico, hace uso de tres constantes de suavización: la primera para estimar el nivel actual de los valores de series de tiempo, la segunda para estimar la pendiente de la línea de tendencia y la tercera para estimar el factor estacional por emplear como multiplicador.

Las cuatro ecuaciones utilizadas en el suavizamiento de Winters son:

1. El nivel de la serie estimado por suavizamiento exponencial:

$$L_t = \alpha \frac{Y_t}{S_{t-s}} + (1 - \alpha)(L_{t-1} + T_{t-1}) \quad (7)$$

2. El estimado de la tendencia:

$$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1} \quad (8)$$

3. El estimado de la estacionalidad:

$$S_t = \gamma \frac{Y_t}{L_t} + (1 - \gamma)S_{t-s} \quad (9)$$

4. Pronóstico del periodo p en el futuro:

$$\hat{Y}_{t+p} = (L_t + pT_t)S_{t-s+p} \quad (10)$$

Donde:

L_t = nuevo valor suavizado (estimado del Nivel actual).

α = constante de suavizamiento para el Nivel ($0 < \alpha < 1$).

Y_t = observación nueva o valor real de la serie en el periodo t .

β = constante de suavizamiento para el estimado de tendencia ($0 < \beta < 1$).

T_t = estimado de tendencia.

γ = constante de suavizamiento para el estimado de estacionalidad. ($0 < \gamma < 1$).

S_t = estimado de estacionalidad.

p = periodos a pronosticarse en el futuro.

s = longitud de la estacionalidad.

\hat{Y}_{t+p} = pronósticos para el periodo p en el futuro.

1.5.2 PROMEDIO MÓVIL

El método más simple para el pronóstico por series de tiempo es el método del promedio móvil. En este método la serie de tiempo presenta aleatoriedad. Un promedio móvil de orden k se calcula mediante:

$$\hat{Y}_{t+1} = \frac{Y_t + Y_{t-1} + \dots + Y_{t-k+1}}{k} \quad (11)$$

Donde

\hat{Y}_{t+1} = valor pronosticado para el siguiente periodo.

Y_t = valor real en el periodo t .

k = número de periodos.

1.6 ERROR DE PRONÓSTICO O RESIDUAL

El error de pronóstico o residual es la resta entre el valor real y su pronóstico. Para calcular el error de cada periodo pronosticado se utiliza la siguiente ecuación:

$$e_t = Y_t - \hat{Y}_t \quad (12)$$

Donde

e_t = error de pronóstico en el periodo t .

Y_t = valor real en el periodo t .

\hat{Y}_t = valor de pronóstico para el tiempo t .

1.7 MEDIDAS DE PRECISIÓN

Un método para evaluar y comparar las técnicas de pronósticos es con la utilización de la suma de errores absolutos, es decir si la técnica posee errores pequeños es la adecuada para generar los pronósticos.

1.7.1 MEAN ABSOLUTE DESVIATION (MAD).

Las siglas en español significan Desviación Media Absoluta, MAD es adecuado cuando las unidades del error van a ser iguales que los valores reales.

$$MAD = \frac{1}{n} \sum_{t=1}^n |Y_t - \hat{Y}_t| \quad (13)$$

Donde:

Y_t = valor real en el periodo t.

\hat{Y}_t = valor de pronóstico para el tiempo t.

t = tiempo.

n = tamaño de la muestra.

1.7.2 MEAN SQUARED ERROR (MSE).

Las siglas en español significan Error Cuadrático Medio, indican si el error de pronóstico es grande o pequeño en comparación con los valores reales.

$$MSE = \frac{1}{n} \sum_{t=1}^n (Y_t - \hat{Y}_t)^2 \quad (14)$$

Donde:

Y_t = valor real en el periodo t.

\hat{Y}_t = valor de pronóstico para el tiempo t.

t = tiempo.

n = tamaño de la muestra

1.7.3 MEAN ABSOLUTE PORCENTAGE ERROR (MAPE).

Las siglas en español significan Error Porcentual Absoluto Medio, se calcula encontrando el error absoluto en cada periodo y promediando los errores porcentuales absolutos, nos indican una comparación entre distintas técnicas de diferentes variables. (2)

$$MAPE = \frac{1}{n} \sum_{t=1}^n \frac{|Y_t - \hat{Y}_t|}{Y_t} \quad (15)$$

Donde:

Y_t =valor real en el periodo t.

\hat{Y}_t =valor de pronóstico para el tiempo t.

t = tiempo.

n = tamaño de la muestra

1.8 PRUEBA DE HIPÓTESIS

Muchos problemas requieren decidir si se acepta o rechaza un enunciado acerca de algún parámetro. El enunciado y el procedimiento de toma de decisiones recibe el nombre de prueba de hipótesis. Este es uno de los aspectos más útiles de la inferencia estadística, puesto que muchos tipos de problemas de decisión pueden formularse y buscar una solución.

Existen dos clases de hipótesis estadísticas:

Hipótesis Nula H_0 , es la afirmación sobre una o más características de poblaciones que al inicio se supone cierta(es decir, la “creencia a priori”).

Hipótesis Alternativa H_1 , es la afirmación contradictoria a H_0 , y esta es la hipótesis del investigador.

1.8.1 ERRORES DE TIPO I Y TIPO II

La decisión para aceptar o rechazar la hipótesis nula se basa en una estadística de prueba calculada a partir de los datos en una muestra aleatoria. Cuando se toma una decisión utilizando la información en una muestra aleatoria, esta decisión está sujeta a error. Puede producirse dos tipos de errores cuando se prueban hipótesis.

		Naturaleza de H_0	
		Verdad	Falsa
Decisión	Aceptar H_0	Decisión correcta	Error del tipo II
	Rechazar H_0	Error del tipo I	Decisión correcta

Si la hipótesis nula se acepta cuando es falsa, entonces el error cometido es del tipo II.

Si la hipótesis nula se rechaza cuando es verdadera, entonces se ha cometido un error de tipo I.

Probabilidad de cometer un error de tipo I se denota por α y se fija al escoger la región de rechazo. Este valor se denomina nivel de significación de la prueba estadística:

$$P(\text{error tipo I}) = \text{nivel de significación} = \alpha$$

Probabilidad de cometer un error de tipo II se denota por β y depende de la hipótesis alternativa que escoja:

$$P(\text{error tipo II}) = \beta$$

1.8.2 PASOS PARA REALIZAR UNA PRUEBA DE HIPÓTESIS

PASO 1. Planteamiento de hipótesis, siempre se parte de una hipótesis nula la que indicará lo que queremos probar, es decir la que suponemos que va a resultar verdadero; y la hipótesis alternativa es la que indicara si mi prueba será unilateral o bilateral.

PASO 2. Selección del nivel de significancia, especificar el nivel de significancia es importante para el análisis por lo general se trabaja con los valores de 0.01, 0.05 ó 0.10.

PASO 3. Descripción del tamaño de la muestra y cálculo del estadístico de prueba, dentro de este paso se detalla el numero de datos que se encuentran en cada variable o en toda la muestra, a su vez es relevante saber de qué distribución provienen los datos para identificar el estadístico de prueba mas adecuado.

El estadístico de prueba permite aceptar o rechazar la hipótesis nula, paso muy importante para la toma de decisiones.

Por ejemplo:

Tabla 1. SELECCIÓN DEL ESTADÍSTICO DE PRUEBA.

Prueba de hipótesis	Estadístico de prueba
Acerca de una media poblacional	t de Student z (distribución normal)
Acerca de la diferencia entre dos medias independientes	t de Student
Acerca de la diferencia entre dos medias dependientes (pareadas)	t de Student
Acerca de la varianza de una población normal	χ^2 (Ji-cuadrada)
Para la comparación entre dos varianzas	F de Fisher
Acerca de una proporción poblacional	z (distribución normal)
Para la comparación entre dos proporciones poblacionales	z (distribución normal)

PASO 4. Especificación de las regiones de aceptación y rechazo, una vez que se estableció si el análisis es unilateral o bilateral por medio de graficas se muestran los valores críticos siendo los limites.

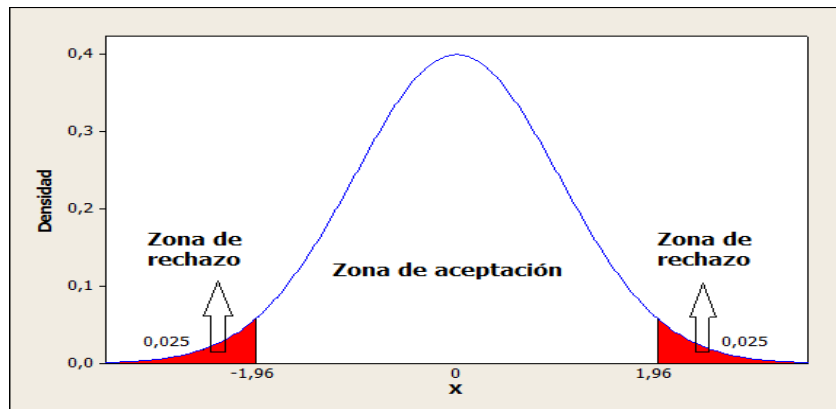


Ilustración 5: ZONA DE ACEPTACIÓN Y RECHAZO DE LA DISTRIBUCIÓN NORMAL $\alpha=0.05$

PASO 5. Decisión estadística, para obtener una decisión correcta hay que tomar en cuenta la comparación del estadístico de prueba con los valores críticos, si el estadístico de prueba está dentro de la zona de aceptación, se dice que se acepta H_0 ; si el estadístico de prueba está en la zona de rechazo, se dice que se acepta H_1 . (5)

CAPÍTULO II

METODOLOGÍA DE LA INVESTIGACIÓN

2.1 LUGAR DE LA INVESTIGACIÓN

La presente investigación se desarrolló en el Departamento de Planificación de la Empresa Eléctrica de la ciudad de Riobamba.

2.2 MÉTODOS Y TÉCNICAS

2.2.1 ETAPAS DEL AED

2.2.1.1 Organización de Datos

Se trabajó con una muestra de 48 datos por variable y se organizó de la siguiente manera:

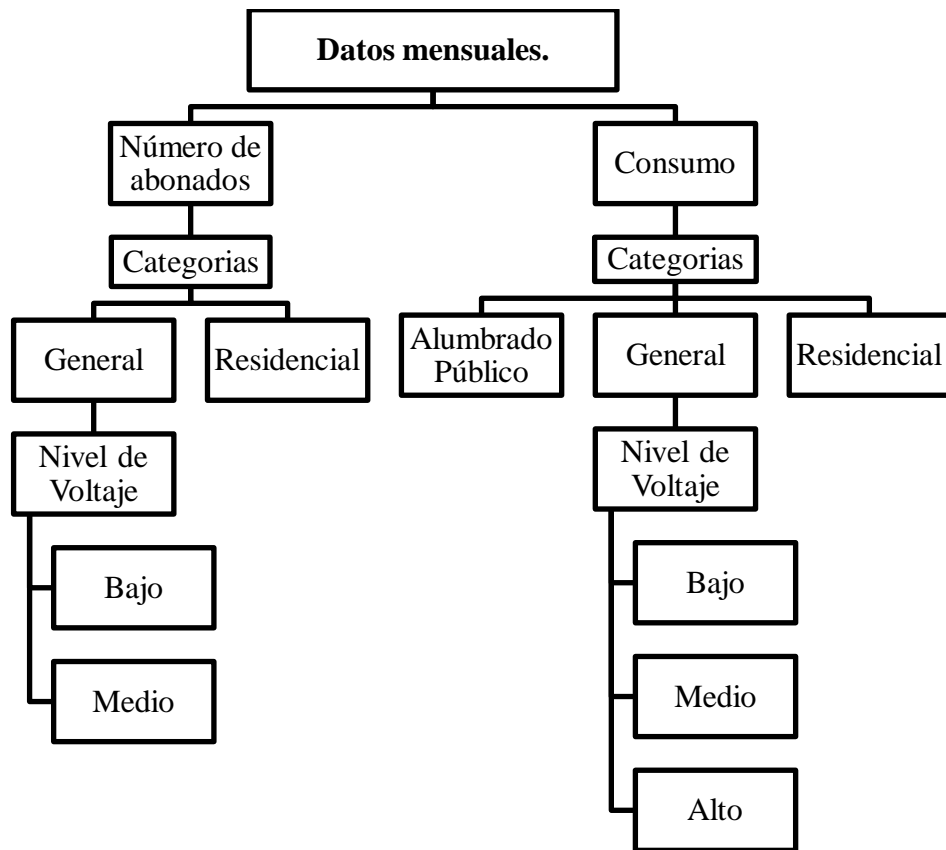


Ilustración 6: ORGANIZACIÓN DE LOS DATOS.

2.2.1.2 Identificación de Datos Atípicos

Para la identificación de datos atípicos se utilizó dos tipos de métodos:

Método Gráfico: se recurrió al BOX-PLOT que es el más indicado para visualizar los datos son atípicos.

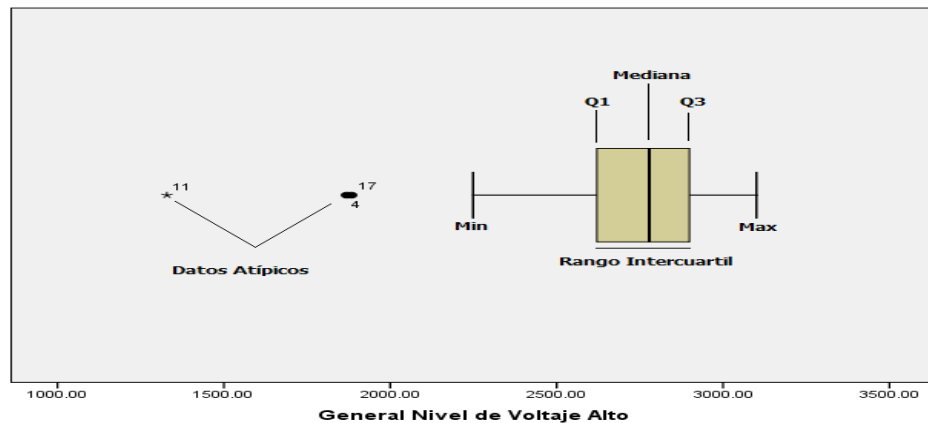


Ilustración 7: MÉTODO GRÁFICO.

Método Analítico: empleamos el método del cuartil para identificar un dato atípico leve y extremo.

Tabla 2. IDENTIFICACIÓN DE DATOS ATÍPICOS LEVES

General Alto	Q1	Q3	IQR	$Q_1 - 1.5 IQR$	$Q_3 + 1.5 IQR$	Comparación	Identificación
1869	2617,87	2901,37	283,5	2192,65	3326,65	$1869 < 2192,65$	Dato atípico

Elaborado por: Valeria Vacacela

Tabla 3. IDENTIFICACIÓN DE DATOS ATÍPICOS EXTREMOS

General Alto	Q1	Q3	IQR	$Q_1 - 3 IQR$	$Q_3 + 3 IQR$	Comparación	Identificación
1869	2617,87	2901,37	283,5	1767,4	3751,9	$1767,4 > 1869 < 3751,9$	No atípico

Elaborado por: Valeria Vacacela

2.2.1.3 Prueba de Normalidad

Una vez identificado los datos atípicos se procede a retirar y poder aplicar la prueba de normalidad, en este caso tenemos que por el tamaño de la muestra se decide aplicar la prueba de Kolmogorov Smirnov, el cual es evaluada por el valor p es decir:

Si el valor $p > 0,05$ = proviene de una distribucion normal

Si el valor $p < 0,05$ = no proviene de una distribucion normal

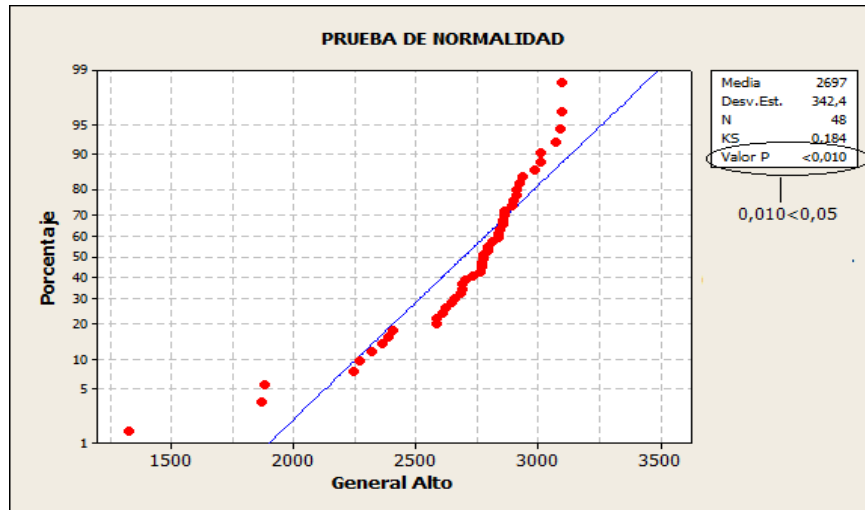


Ilustración 8: PRUEBA DE NORMALIDAD

Existe una variable que no fue proveniente de una distribución normal, aquella variable fue transformada por medio de el logaritmo natural (\ln).

2.2.2 ELECCIÓN DE UNA TÉCNICA DE PRONÓSTICO

A fin de seleccionar adecuadamente la técnica conveniente de pronósticos, se realizó lo siguiente:

- Exploración de las componentes de las series temporales por medio de gráficas de series de tiempo y autocorrelogramas.
- Evaluar los modelos por medio de las medidas de precisión (MAPE, MAD, MSE).
- Evaluamos los modelos utilizando autocorrelogramas para identificar si existe aleatoriedad.
- Utilizamos los modelos para la generación de pronósticos.

Un factor importante que influye en la selección de una técnica de pronóstico es identificar y entender los patrones históricos de los datos

2.2.3 PRUEBA DE HIPÓTESIS

PASO 1. Formulación de hipótesis

$$H_0: \mu_{reales} = \mu_{pronóstico}$$

$$H_1: \mu_{reales} \neq \mu_{pronóstico}$$

PASO 2. Selección de un nivel de significancia

$$\alpha = 0,05$$

PASO 3. Descripción del tamaño de la muestra y cálculo del estadístico de prueba

Tenemos una muestra de $n=8$ por cada variable, debido que en los últimos 4 meses del año 2013 se generaron perdidas y fueron aisladas del análisis.

PASO 4. Especificación de las regiones de aceptación y rechazo

Como se tiene una muestra pequeña se aplica la prueba t–student.

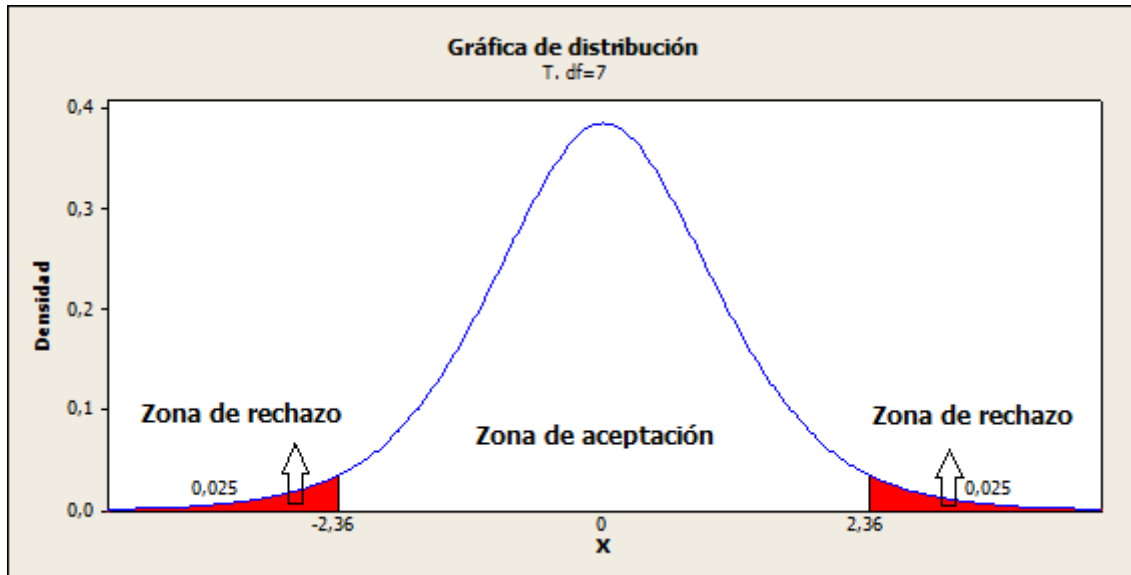


Ilustración 9: REGIONES DE ACEPTACIÓN Y RECHAZO DE LA PRUEBA T-STUDENT

PASO 5. Decisión estadística, si el estadístico de prueba está dentro de la zona de aceptación, se dice que se acepta H_0 ; si el estadístico de prueba está en la zona de rechazo, se dice que se acepta H_1 .

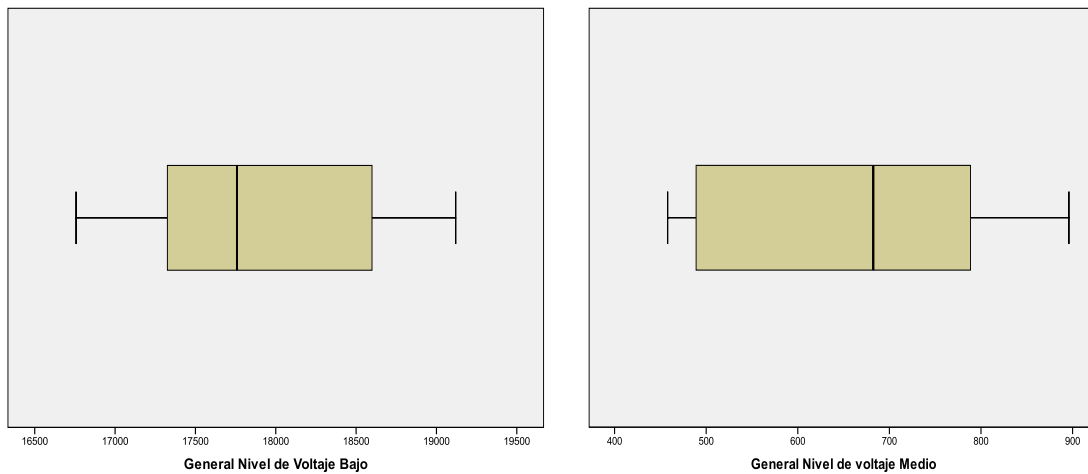
CAPÍTULO III

ANÁLISIS DE DATOS Y RESULTADOS

3.1 IDENTIFICACIÓN DE POSIBLES DATOS ATÍPICOS EN EL PERIODO 2009-2012

3.1.1 MÉTODO GRÁFICO

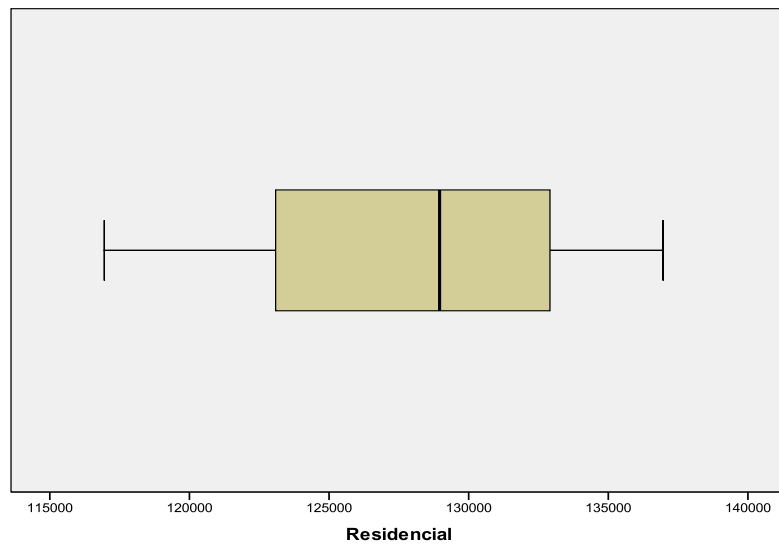
3.1.1.1 BOX-PLOT Número de Abonados



Gráfica 1. NÚMERO DE ABONADOS CATEGORÍA GENERAL EN EL NIVEL DE VOLTAJE BAJO Y MEDIO.

La gráfica 1 muestra la ausencia de datos atípicos en el número de abonados en la categoría general en el nivel de voltaje bajo y medio, en la categoría general en el nivel de voltaje bajo es asimétrica a la derecha con una mediana entre 18000 y 17500 abonados, la variable en la

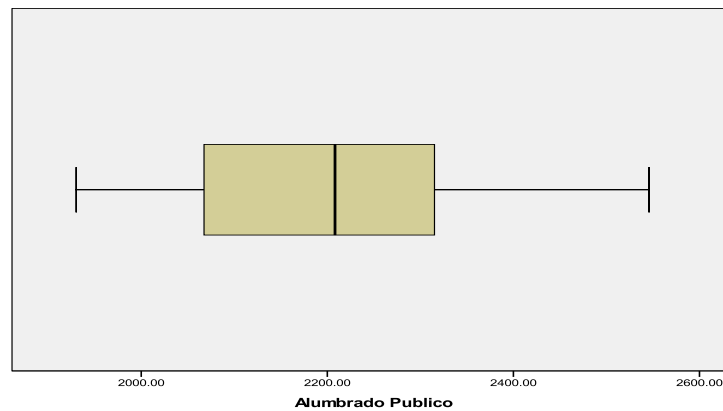
categoría general en el nivel de voltaje medio es asimétrica a la izquierda tiene una mediana que está entre 700 y 600 abonados.



Gráfica 2. NÚMERO DE ABONADOS CATEGORÍA RESIDENCIAL.

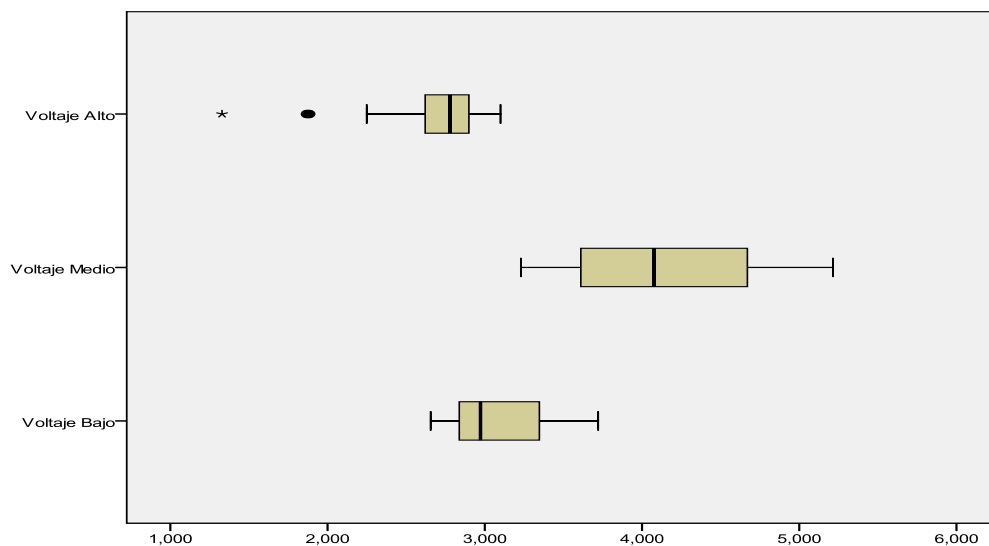
La variable en la categoría residencial es asimétrica a izquierda su mediana está entre 130000 y 125000 abonados, no se identifica la presencia de datos atípicos.

3.1.1.2 BOX-PLOT del Consumo de Energía en MWh



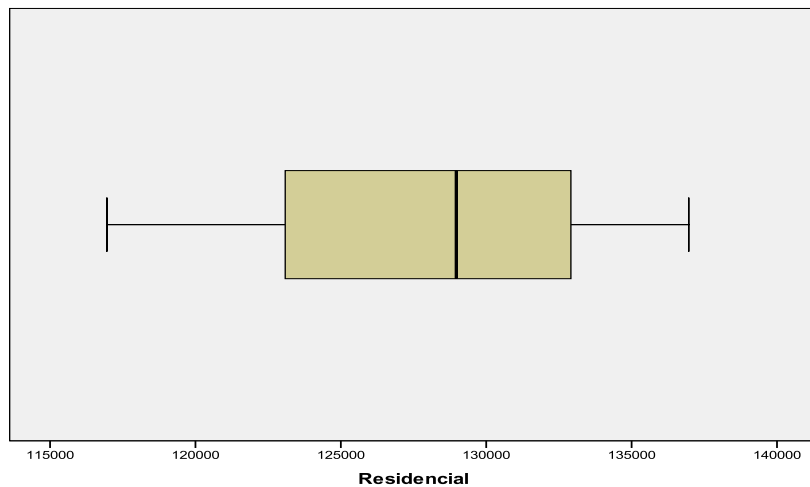
Gráfica 3. CONSUMO DE ENERGÍA EN MWh EN LA CATEGORÍA ALUMBRADO PÚBLICO.

La gráfica 3 indica que en la variable de Alumbrado Público su mediana está cerca de los 2200 MWh, la distribución del consumo es asimétrica a la izquierda.



Gráfica 4. CONSUMO DE ENERGÍA EN MWh DE LA CATEGORIA GENERAL

Se realizó una comparación entre las variables de la categoría general, en el nivel bajo y medio existe una asimetría a la derecha y en el nivel alto tiene una asimetría a la izquierda y a su vez se detecta la presencia de datos atípicos, también se visualiza una diferencia del consumo de energía entre los niveles de la categoría general.



Gráfica 5. CONSUMO DE ENERGÍA EN MWh DE LA CATEGORÍA RESIDENCIAL.

En la variable residencial la mediana se encuentra entre 8000 y 8500 MWh, es asimétrica a la derecha.

3.1.2 MÉTODO ANALÍTICO

Mediante el método de cuartiles se identificó datos atípicos en la variable consumo de energía categoría general nivel de voltaje alto.

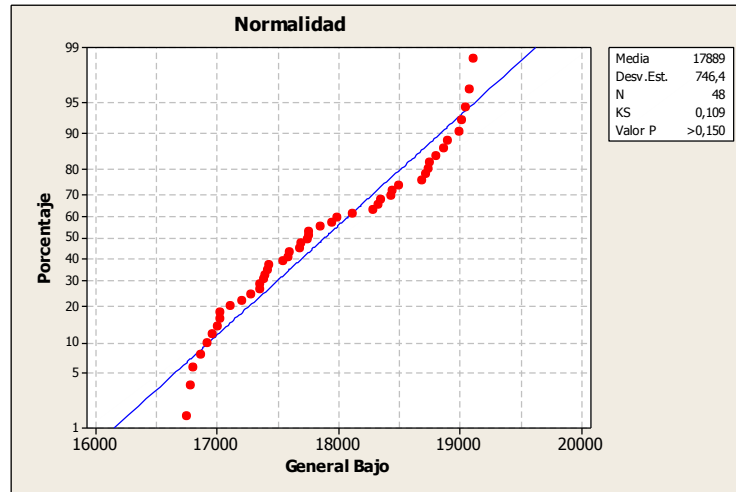
Valores atípicos leves son: 1869, 1329, 1883.

Valores atípicos extremos son: 1329, 1883.

3.2 DEPURACIÓN DE DATOS Y PRUEBA DE NORMALIDAD

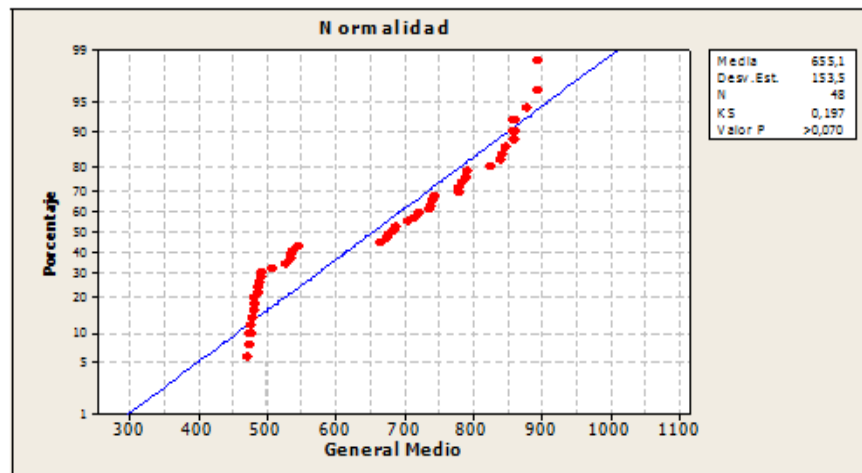
Los datos identificados como atípicos han sido excluidos del análisis.

3.2.1 PRUEBA DE KOLMOGOROV SMIRNOV PARA EL NUMERO DE ABONADOS



Gráfica 6. PRUEBA DE NORMALIDAD PARA EL NACGNVB

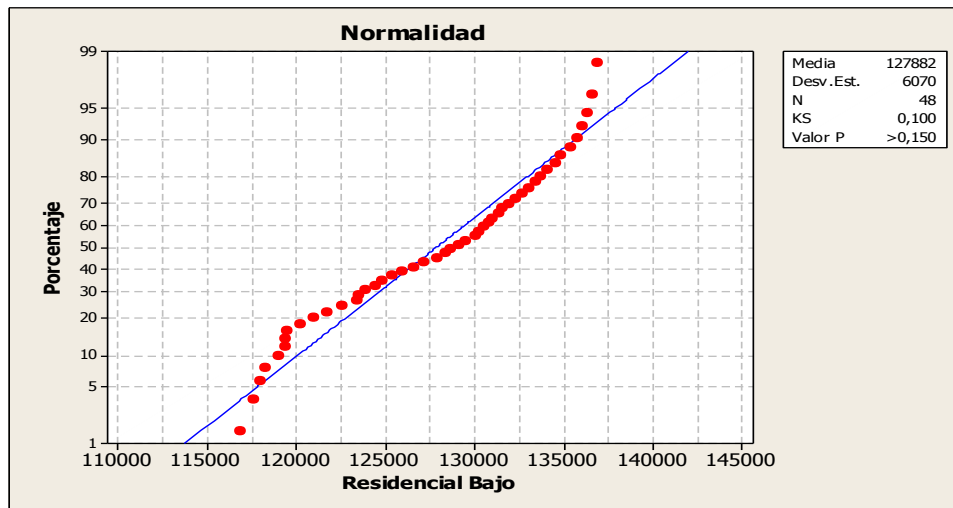
En la gráfica 6 el número de abonados en la categoría general en el nivel de voltaje bajo se ajusta a una recta obteniendo así un valor p de 0.15 siendo mayor al nivel de significancia de 0.05; por lo tanto el número de abonados de la categoría general en el nivel de voltaje bajo se ajustan a una distribución normal.



Gráfica 7. PRUEBA DE NORMALIDAD PARA EL NACGNVM

La gráfica 7 indica que los datos del número de abonados se encuentran cercanos a la línea de normalidad, tomando en cuenta que el valor p de 0,07 es mayor al nivel de significancia

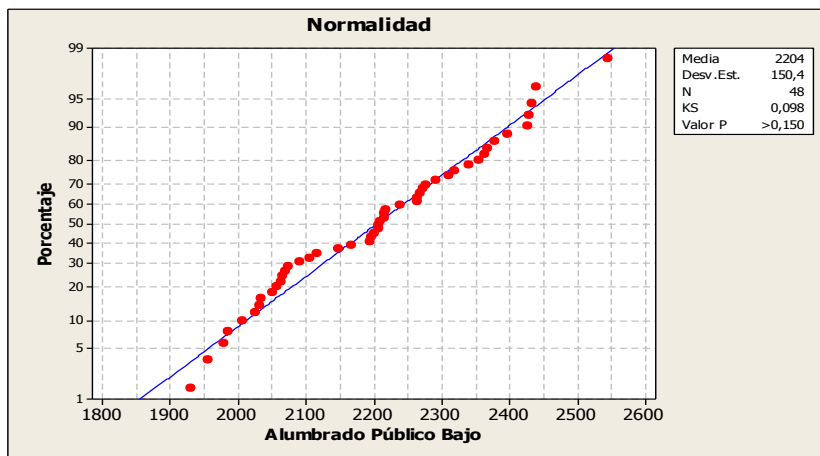
de 0,05 lo que muestra que la variable proviene de una distribución normal.



Gráfica 8. PRUEBA DE NORMALIDAD PARA EL NACR

Los datos se ajustan a una recta obteniendo así un valor p de 0,15 siendo mayor al nivel de significancia, comprobamos que la información proviene de una distribución normal.

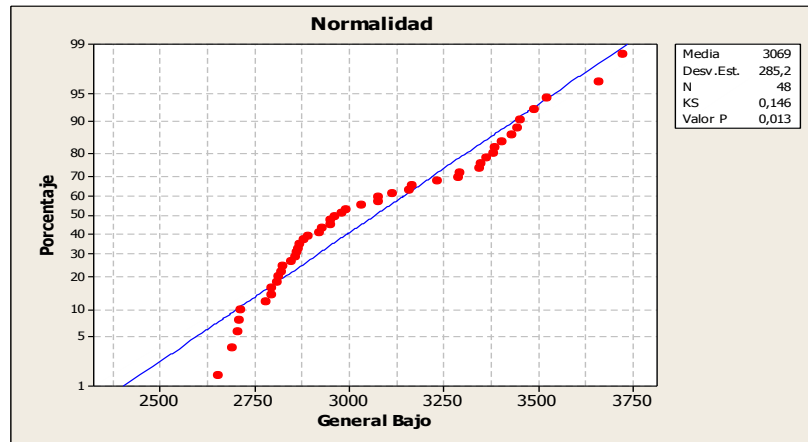
3.2.2 PRUEBA DE KOLMOGOROV SMIRNOV PARA EL CONSUMO DE ENERGÍA EN MWh



Gráfica 9. PRUEBA DE NORMALIDAD PARA EL CECAP

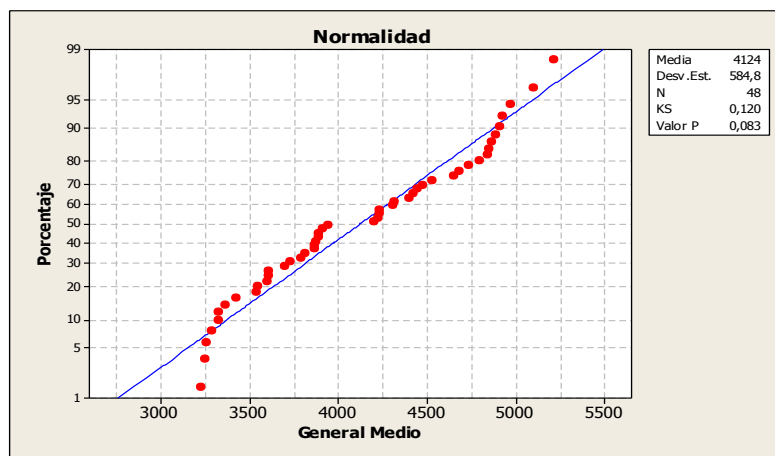
Los datos que se observan en la gráfica 9 muestran que existen puntos cerca y sobre la línea,

nos entrega un valor p del 0,15 que es mayor al nivel de significancia del 0,05, por lo tanto los datos provienen de una distribución normal.



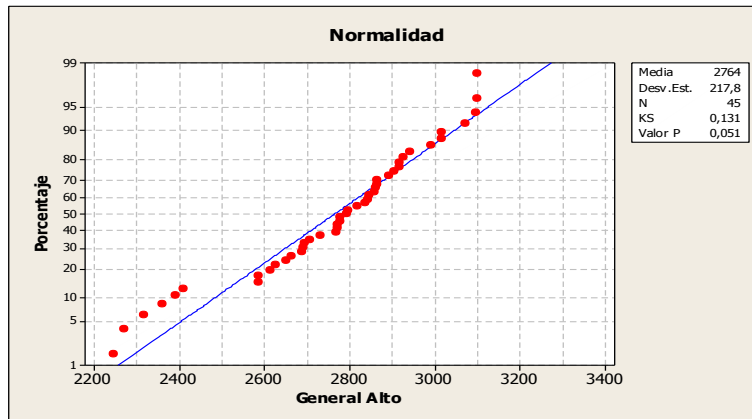
Gráfica 10. PRUEBA DE NORMALIDAD PARA EL CECGNVB

El consumo de energía en MWH en la categoría general, voltaje bajo no proviene de una distribución normal ya que el valor $p = 0,013$ y $\alpha = 0,05$ por lo tanto $p < \alpha$.



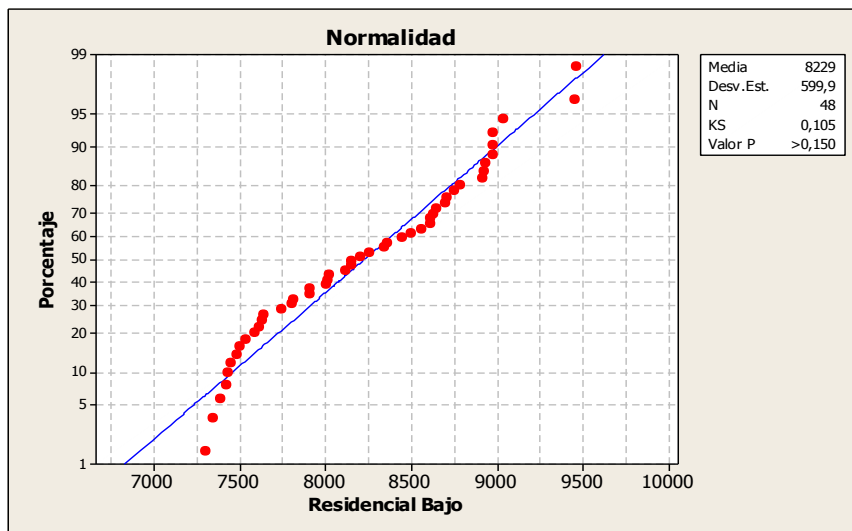
Gráfica 11. PRUEBA DE NORMALIDAD PARA EL CECGNVM

Los puntos que se visualizan en la gráfica 11 están cercanos a la línea, son muy pocos los puntos que están lejos, el valor p que se obtuvo es de 0,08 el cual es mayor al nivel de significancia, por lo tanto los datos consumo de energía se ajustan a una distribución normal.



Gráfica 12. PRUEBA DE NORMALIDAD PARA EL CECGNA

Los datos del consumo de energía de la grafica 12 están cercanos a la línea y tiene un valor p de 0,051 el cual es mayor al nivel de significancia de 0,05 indicando que los datos provienen de una distribución normal.



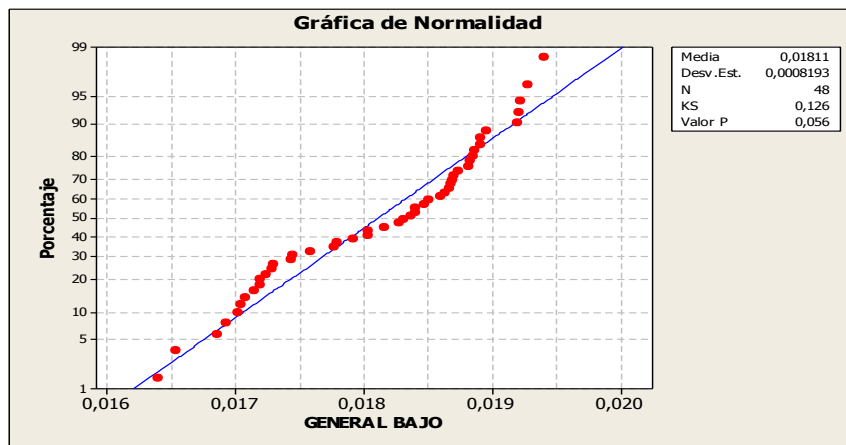
Gráfica 13. PRUEBA DE NORMALIDAD PARA EL CECR

En la gráfica 13 los datos se ajusta a una recta obteniendo así un valor p de 0.15 siendo mayor al nivel de significancia de 0.05; por lo tanto el consumo de energía proviene a una distribución normal.

3.3 TRANSFORMACION DE LA VARIABLE QUE NO SE AJUSTA A LA NORMAL

La prueba de normalidad indicó en la grafica 8 que la variable consumo en la categoría general en el nivel de voltaje bajo no se ajusta a una distribución normal por lo tanto se requiere hacer una transformación de los datos por medio de logaritmo natural.

Se aplicó la prueba de Kolmogorov Smirnov en los datos transformados y se muestra la siguiente grafica:

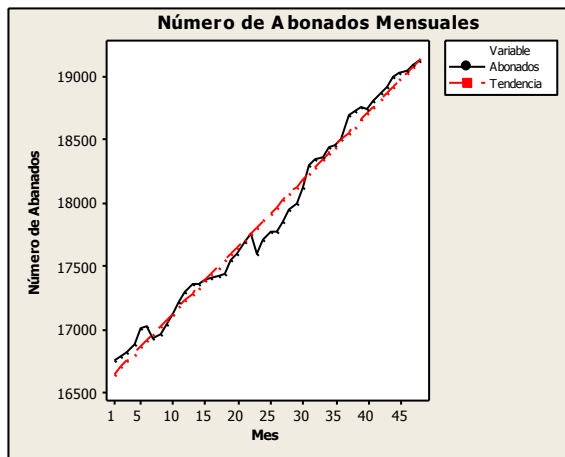


Gráfica 14. NORMALIDAD DE DATOS TRANSFORMADOS DEL CECGNV

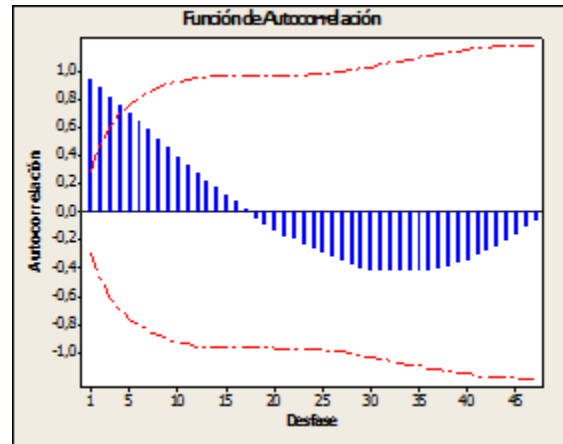
Se obtuvo un valor p de 0,056 el cual supera a 0,05 es decir que los datos transformados provienen de una distribución normal.

3.4 EXPLORACIÓN DE LAS COMPONENTES DE LAS SERIES TEMPORALES.

3.4.1 NÚMERO DE ABONADOS EN EL PERIODO 2009-2012



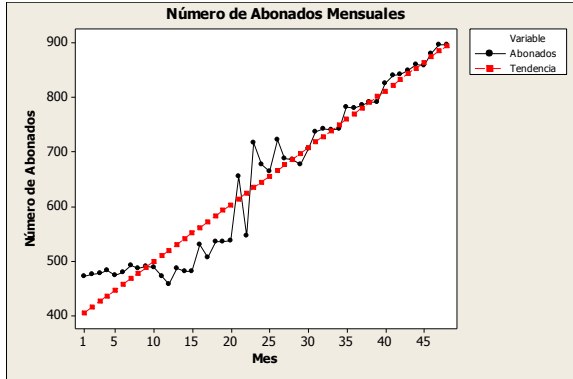
Gráfica 15. NACGNVB



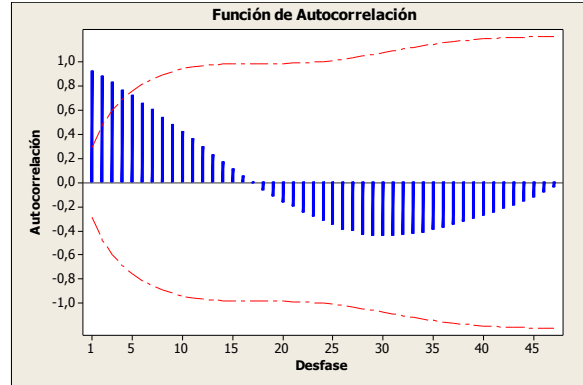
Gráfica 16. FUNCIÓN DE AUTOCORRELACIÓN
PARA EL NACGNVB

La serie representada en la gráfica 15 presentan los datos por arriba y debajo de la línea de tendencia, lo que indica que existe ciclicidad. También se puede concluir que la serie presenta una tendencia.

La gráfica 16 de función de autocorrelación corrobora la existencia de tendencia.

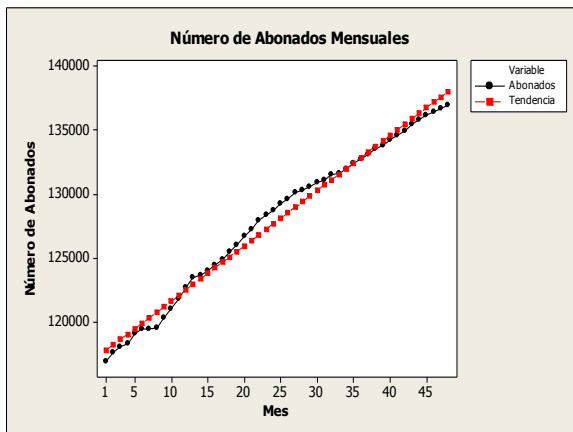


Gráfica 17. NACGNVM

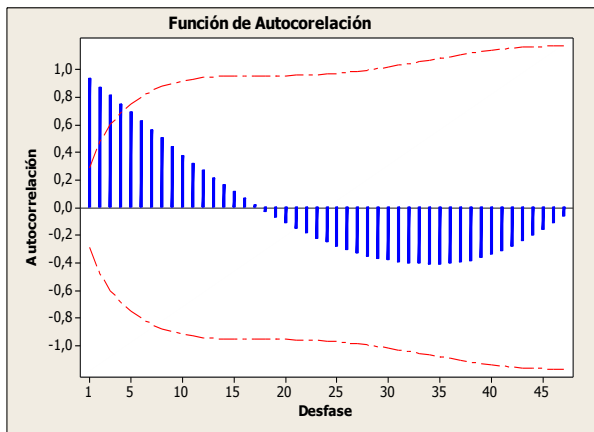


**Gráfica 18. FUNCIÓN DE AUTOCORRELACIÓN
PARA EL NACGNVM**

En la gráfica 17 los datos del número de abonados están por encima y debajo de la línea de tendencia mostrando la existencia de ciclicidad, también se observa una tendencia lo que nos corrobora en la grafica18.



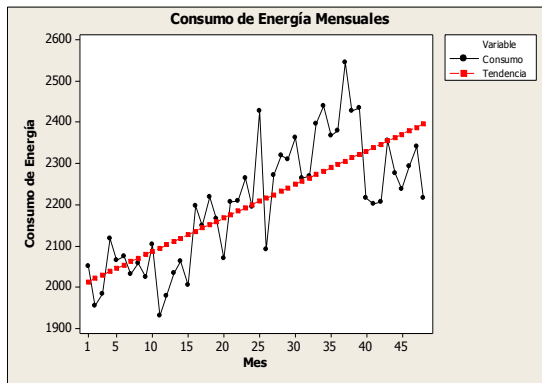
Gráfica 19. NACR



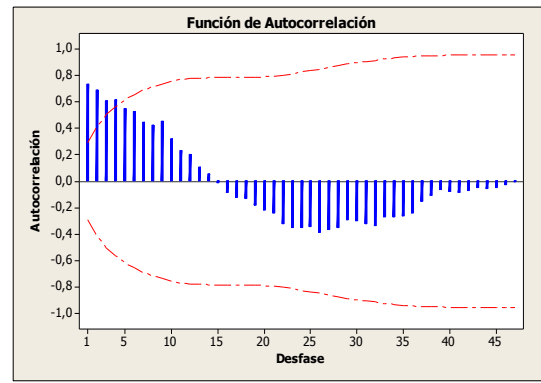
**Gráfica 20. FUNCIÓN DE AUTOCORRELACIÓN
PARA EL NACR**

El número de abonados que se visualiza en la gráfica 19 indica que se encuentran sobre línea de tendencia, por lo que es evidente la existencia de las componentes de ciclicidad y tendencia, la grafica 20 corrobora la existencia de tendencia.

3.4.2 GRÁFICAS DE SERIES DE TIEMPO Y AUTOCORRELOGRAMAS PARA DATOS DEL CONSUMO DE ENERGÍA EN MWh EN EL PERIODO 2009-2012

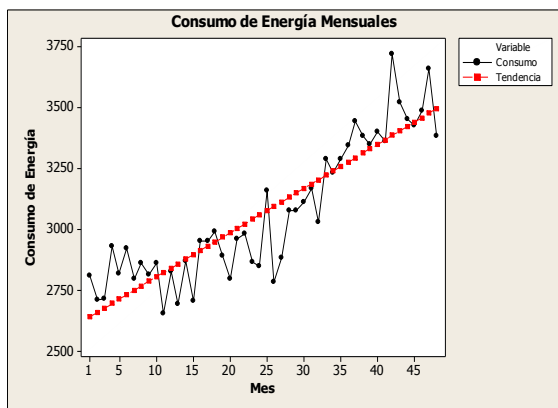


Gráfica 21. SERIES DE TIEMPO CECAP

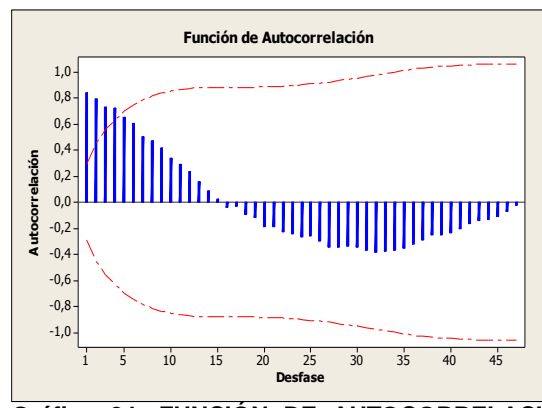


Gráfica 22. FUNCIÓN DE AUTOCORRELACIÓN PARA EL CECAP

Los datos representados en la gráfica 21 muestran datos por arriba y debajo de la línea de tendencia. La gráfica 22 nos ayuda a verificar la presencia de tendencia.

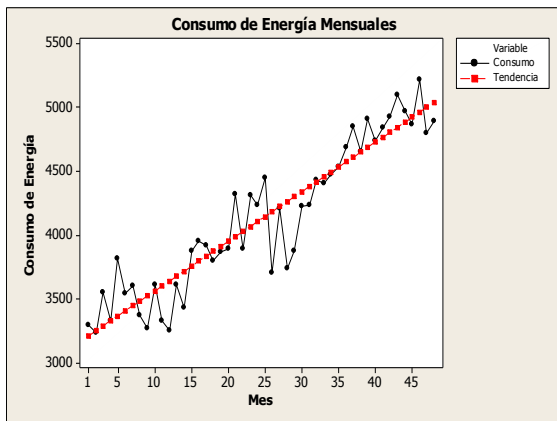


Gráfica 23. SERIES DE TIEMPO CECGNVB

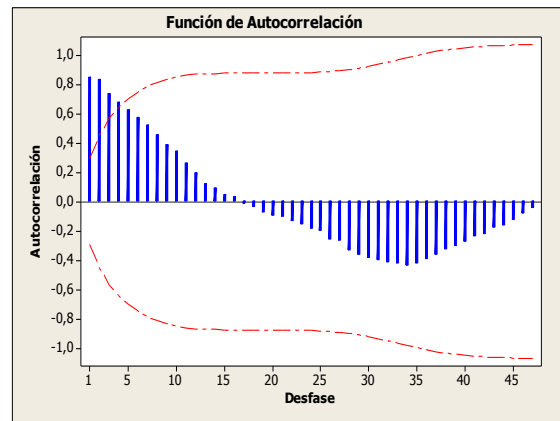


Gráfica 24. FUNCIÓN DE AUTOCORRELACIÓN PARA EL CECGNVB

Los datos de consumo en la gráfica 23 indican la presencia de tendencia y ciclicidad debido a que los datos se encuentran por encima y debajo de la línea de tendencia. En la gráfica 24 indica la existencia de tendencia.

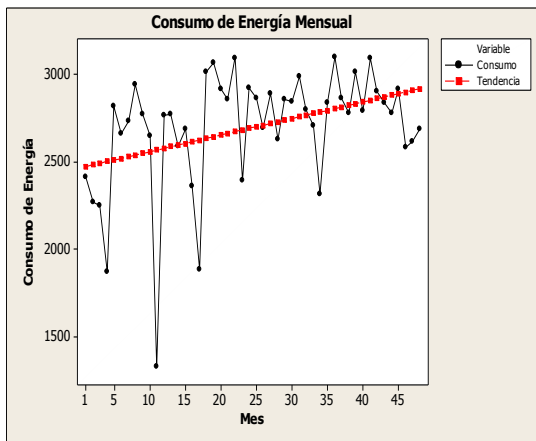


Gráfica 25. SERIES DE TIEMPO CECGNVM

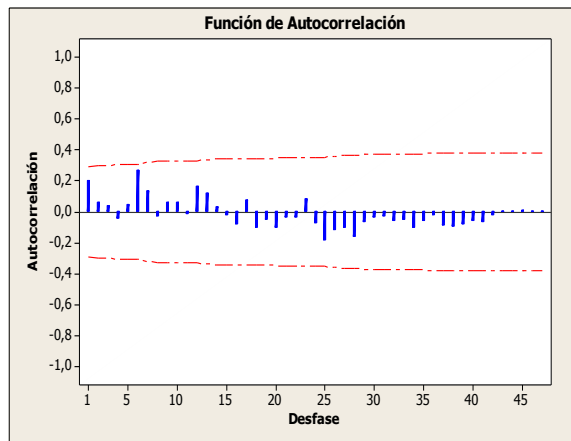


Gráfica 26. FUNCIÓN DE AUTOCORRELACIÓN PARA EL CECGNVM

La serie representada en la gráfica 25 indica que el consumo de energía se encuentra por encima y debajo de la línea de tendencia por lo tanto hay la presencia de ciclicidad y tendencia, con la función de autocorrelación en la gráfica 26 corrobora la tendencia



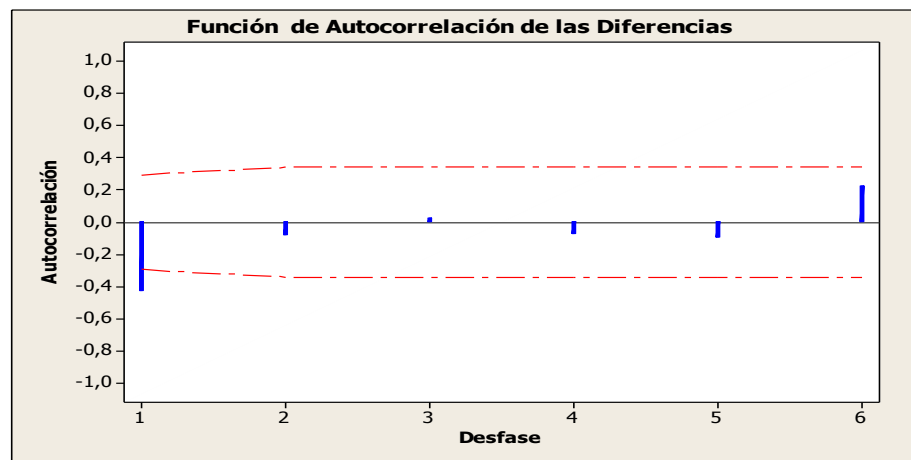
Gráfica 27. SERIES DE TIEMPO CECGNVA



Gráfica 28. FUNCIÓN DE AUTOCORRELACIÓN PARA EL CECGNVA

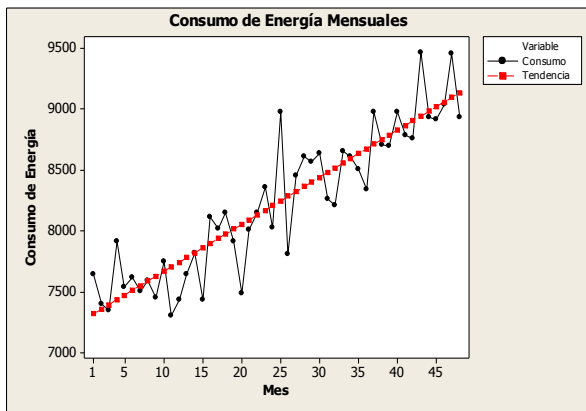
Observando la gráfica 27 nos indica que existe ciclicidad también se puede mirar que el consumo de energía no presenta con claridad una tendencia. La gráfica 28 de función de autocorrelación corrobora la falta de tendencia y la existencia de aleatoriedad.

Debido a la presencia de aleatoriedad es necesario realizar diferencias del consumo de energía en MWh de la categoría general en el nivel de voltaje alto en el periodo 2009-2012 y a su vez representar en una gráfica de función de autocorrelación con un desfase.

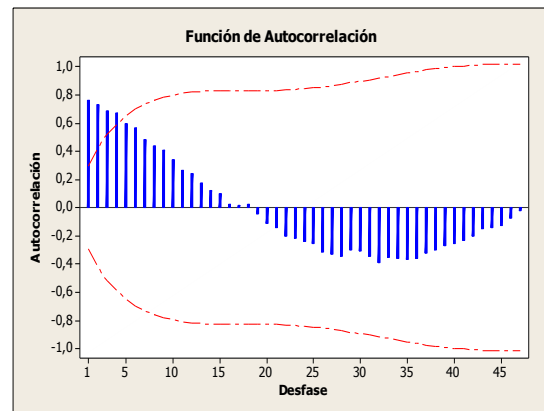


Gráfica 29. FUNCIÓN DE AUTOCORRELACIÓN DE LAS DIFERENCIAS PARA EL CECGNVA

La gráfica 29 de función de autocorrelación de las diferencias para el consumo de energía en MWh de la categoría general en el nivel de voltaje alto en el periodo 2009-2012 nos indica que la componente subyacente es la estacionariedad, esto nos confirma que la componente estacionalidad no existe.



Gráfica 30. SERIES DE TIEMPO CECR



Gráfica 31. FUNCIÓN DE AUTOCORRELACIÓN PARA EL CECR

La serie representada en la gráfica 30 presentan datos por arriba y debajo de la línea de tendencia, lo que indica que existe ciclicidad. La gráfica 31 de función de autocorrelación corrobora la existencia de tendencia.

3.5 IDENTIFICACIÓN DEL MODELO DE PRONÓSTICO PARA DATOS DEL PERIODO 2009-2012

3.5.1 MODELOS PARA DATOS DEL NÚMERO DE ABONADOS

Tabla 4. MODELOS PARA EL NACGNVB

Modelo	Medidas de Precisión		
	MAPE	MAD	MSE
Lineal	0,40	70,62	7841,88
Curva S	0,39	69,98	7276,08
Curva de crecimiento	0,37	65,7	6652,88
Cuadrática	0,32	58,18	5195,78
Holt	0,24	41,94	3301,72

Elaborado por: Valeria Vacacela

En la tabla 4 se puede visualizar las medidas de precisión de los diferentes modelos para el número de abonados de la categoría general en el nivel de voltaje bajo el modelo de Holt posee las medidas de precisión más pequeñas en comparación de los cuatro modelos

tentativos restantes, por lo tanto el modelo de Holt será un buen pronosticador para dicha variable.

Tabla 5. MODELOS PARA EL NACGNVM

Medidas de Precisión			
Modelos	MAPE	MAD	MSE
Lineal	5,11	28,16	1397,45
Curva S	4,76	27,37	1221,68
Curva de crecimiento	4,76	28,02	1247,03
Cuadrática	4,65	27,18	1207,71
Holt	2,934	18,205	956,785

Elaborado por: Valeria Vacacela

Las medidas de precisión que se exponen en la tabla 5 indica que el modelo Holt posee las medidas más pequeñas por lo tanto será el modelo adecuado para pronosticar el número de abonados de la categoría general en el nivel de voltaje medio para el periodo 2013.

Tabla 6. MODELOS PARA EL NACR

Medidas de Precisión			
Modelo	MAPE	MAD	MSE
Lineal	0	578	449809
Curva S	0	323	152274
Curva de crecimiento	1	684	612591
Cuadrática	0	276	112999
Holt	0,1	157,5	41618,8

Elaborado por: Valeria Vacacela

La tabla 6 indica las medidas de precisión de los diferentes modelos el de Holt posee las medidas de precisión MAD y MSE mas pequeñas en comparación de los cuatro modelos tentativos restantes, por lo tanto el modelo de Holt será un buen pronosticador para para el número de abonados de la categoría residencial en el nivel de voltaje bajo.

3.5.2 MODELOS PARA DATOS DEL CONSUMO DE ENERGÍA EN MWh

Tabla 7. MODELOS PARA EL CECAP

Medidas de Precisión			
Modelo	MAPE	MAD	MSE
Lineal	3,56	79,19	9491,53
Curva S	3,43	75,73	8738,66
Curva de crecimiento	3,61	80,48	9700,03
Cuadrática	3,25	71,87	8129,49
Holt	3,46	76,23	9707,03

Elaborado por: Valeria Vacacela

La tabla 7 indica que las medidas de precisión más pequeñas son las del modelo cuadrático siendo así el mejor modelo para el consumo de energía de la categoría alumbrado público en el nivel.

Tabla 8. MODELOS PARA EL CECGNVB

Medidas de Precisión			
Modelo	MAPE	MAD	MSE
Lineal	3,2	97,4	16217,9
Curva S	3,2	100,8	17057,1
Curva de crecimiento	3,1	93,2	15107,4
Cuadrática	2,7	83,7	1148,5
Holt	3,3	100,5	14844,2

Elaborado por: Valeria Vacacela

Los diferentes modelos que están en la tabla 8 indican que las mejores medidas de precisión son las del modelo cuadrático el cual va a generar un buen pronóstico del consumo de energía en MWh.

Tabla 9. MODELOS PARA EL CECGNVM

Medidas de Precisión			
Modelo	MAPE	MAD	MSE
Lineal	4,1	162,3	44427,7
Curva S	4	159,4	44027,7
Curva de crecimiento	4	160	42672,8
Cuadrática	4	158,7	42359,8
Holt	4,7	186,8	53908,8

Elaborado por: Valeria Vacacela

En la tabla 9 se puede observar que el modelo cuadrático tiene las medidas de precisión mas pequeñas por lo tanto dará un pronóstico adecuado para el consumo de energía en MWh en la categoría de general en el nivel de voltaje medio.

Tabla 10. MODELOS PARA EL CECGNVA

Medidas de Precisión			
Modelos	MAPE	MAD	MSE
Promedio Móvil	12	287	185793
Modelo de Brown	10	250	109600

Elaborado por: Valeria Vacacela

En la tabla 10 las medidas de precisión de los modelos indica que el modelo de Brown tiene las medidas de precisión más pequeñas por lo tanto será el mejor pronostico para el consumo de energía en MWh en la categoría de general en el nivel de voltaje medio.

Tabla 11. MODELOS PARA EL CECR

Medidas de Precisión			
Modelos	MAPE	MAD	MSE
Lineal	9,7	226,4	97504,6
Curva S	9,7	226,5	99986,1
Curva de crecimiento	9,9	234,3	98924,8
Cuadrática	9	207,2	91280,1
Holt	3	242,9	90380,5

Elaborado por: Valeria Vacacela

La tabla 11 indica que el modelo de Holt posee las medidas de precisión más pequeñas dando la seguridad de aplicar para generar los pronósticos del consumo de energía en MWh en la categoría residencial.

3.6 ESTIMACIÓN DE LOS PARÁMETROS DEL LOS MODELOS

Modelo de Holt

Estimado del Nivel actual de la serie suavizada exponencialmente:

$$L_t = \alpha Y_t + (1 - \alpha)(L_{t-1} + T_{t-1})$$

Estimado de tendencia:

$$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1}$$

Pronóstico del periodo p en el futuro:

$$\hat{Y}_{t+p} = L_t + pT_t$$

Donde:

L_t = nuevo valor suavizado

α = constante de suavizamiento

Y_t = observacion nueva o valor real

β = constante de suavizamiento para la tendencia

T_t = estimado de tendencia

p = periodo a pronosticarse en el futuro

\hat{Y}_{t+p} = pronóstico para el periodo p en el futuro

VARIABLE	ESTIMACIÓN
Número de Abonados Categoría General Nivel de Voltaje Bajo.	$L_t = 1,12Y_t + (1 - 1,12)(L_{t-1} + T_{t-1})$ $T_t = 0,02(L_t - L_{t-1}) + (1 - 0,02)T_{t-1}$
Número de Abonados Categoría General Nivel de Voltaje Medio.	$L_t = 0,42Y_t + (1 - 0,42)(L_{t-1} + T_{t-1})$ $T_t = 0,05(L_t - L_{t-1}) + (1 - 0,05)T_{t-1}$
Número de Abonados Categoría Residencial.	$L_t = 1,30Y_t + (1 - 1,30)(L_{t-1} + T_{t-1})$ $T_t = 0,07(L_t - L_{t-1}) + (1 - 0,07)T_{t-1}$
Consumo de Energía Categoría Residencial.	$L_t = 0,41Y_t + (1 - 0,41)(L_{t-1} + T_{t-1})$ $T_t = 0,05(L_t - L_{t-1}) + (1 - 0,05)T_{t-1}$

Modelo Cuadrático.

$$\hat{Y}_t = \beta_0 + \beta_1 t + \beta_2 t^2$$

Donde:

\hat{Y}_t = pronóstico para el periodo t

β_0 = observacion de origen

β_1 = nueva observación en el periodo 1

β_2 = nueva observación en el periodo 1

t = tiempo

VARIABLE	ESTIMACIÓN
Consumo de Energía Categoría Alumbrado Público.	$\hat{Y}_t = 1917,5 + 18,67t - 0,22t^2$
Consumo de Energía Categoría General Nivel de Voltaje Bajo.	$\hat{Y}_t = 2787,4 - 1,47t + 0,40t^2$
Consumo de Energía Categoría General Nivel de Voltaje Media	$\hat{Y}_t = 3278,8 + 25,91t + 0,27t^2$

Modelo de Brown

$$\hat{Y}_{t+1} = \alpha Y_t + (1 + \alpha)\hat{Y}_t$$

Donde:

\hat{Y}_{t+1} = *valor del pronostico para el siguiente periodo*

α = *constante de suavizamiento*

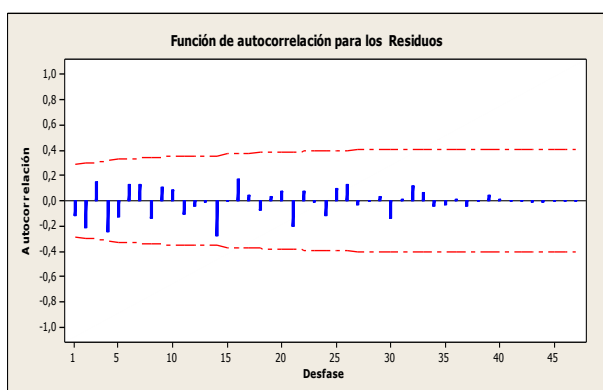
Y_t = *nueva observacion o valor real*

\hat{Y}_t = *antiguo valor suavizado*

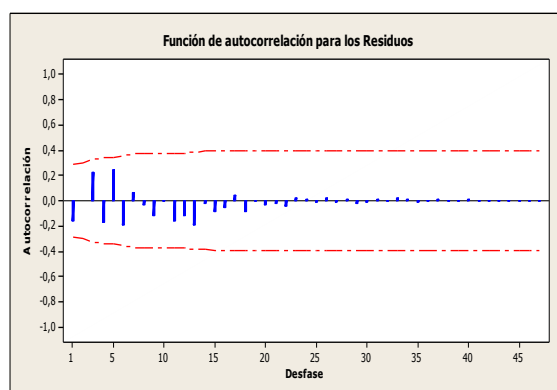
VARIABLE	ESTIMACIÓN
Consumo de Energía Categoría General Nivel de Voltaje Alto.	$\hat{Y}_{t+1} = 0,12Y_t + (1 + 0,12)\hat{Y}_t$

3.7 VALUACIÓN DEL MODELO

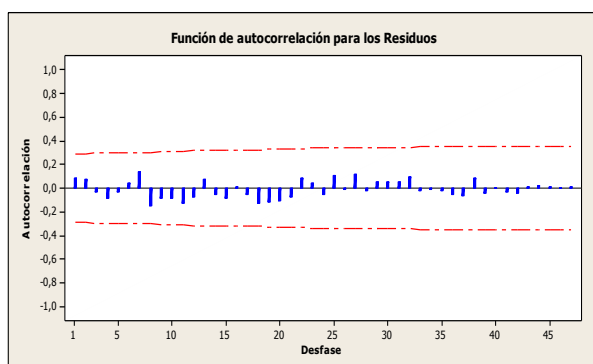
3.7.1 EVALUACIÓN DEL MODELO PARA DATOS DEL NÚMERO DE ABONADOS



Gráfica 32. FUNCIÓN DE AUTOCORRELACIÓN PARA LOS RESIDUOS DEL NACGNVB



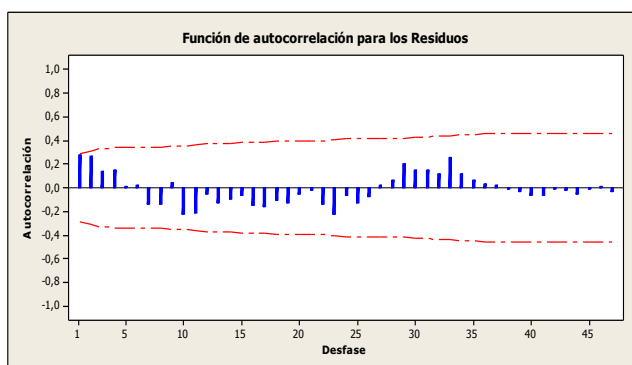
Gráfica 33. FUNCIÓN DE AUTOCORRELACIÓN PARA LOS RESIDUOS DEL NACGNVM



Gráfica 34. FUNCIÓN DE AUTOCORRELACIÓN PARA LOS RESIDUOS DEL NACR

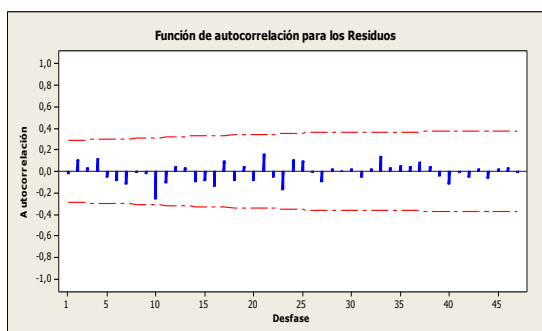
Las gráficas 32, 33 y 34 de autocorrelación del número de abonados en las categorías general en el nivel de voltaje bajo, medio y residencial muestran que los residuos son aleatorios, requisito muy indispensable para la aplicación de los modelos.

3.7.2 EVALUACIÓN DEL MODELO PARA DATOS DEL CONSUMO DE ENERGÍA EN MWH

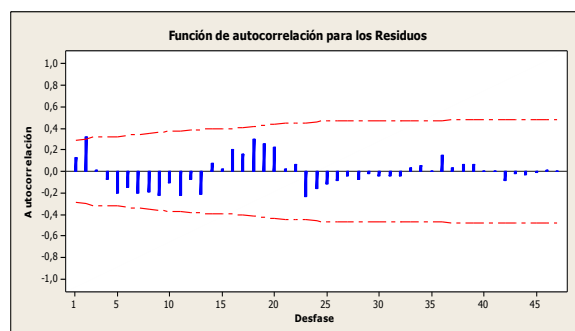


Gráfica 35. FUNCIÓN DE AUTOCORRELACIÓN PARA LOS RESIDUO DEL CECAP

Los residuos del consumo de energía en la categoría alumbrado público indican que existe aleatoriedad, lo que indica que el modelo que se utilizó es el correcto.



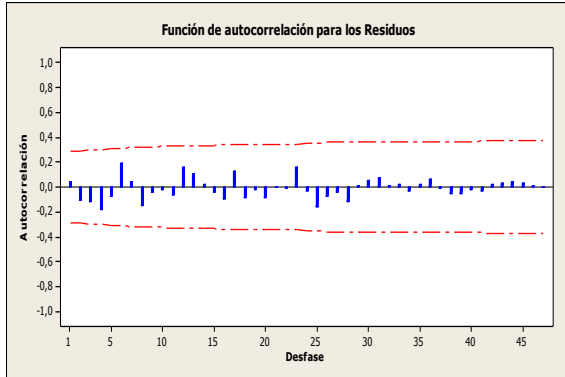
Gráfica 36. FUNCIÓN DE AUTOCORRELACIÓN PARA LOS RESIDUO DEL CECGNVB



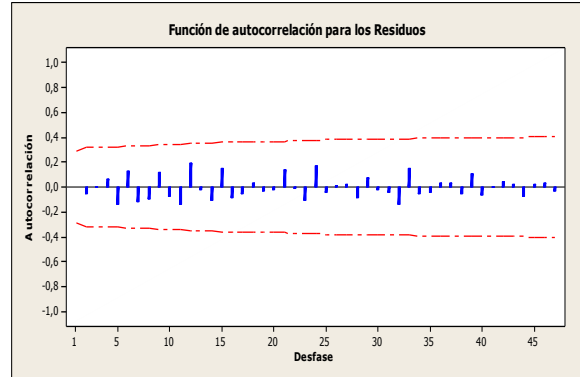
Gráfica 37. FUNCIÓN DE AUTOCORRELACIÓN PARA LOS RESIDUO DEL CECGNVM

En gráfica 36 los residuos del consumo de energía nos indican que son aleatorios, por lo tanto los pronósticos que se generaron con el respectivo modelo son los mejores.

Los residuos de la gráfica 37 indican existe aleatoriedad y por ende el modelo que se utilizó para generar pronósticos es el adecuado.



Gráfica 38. FUNCIÓN DE AUTOCORRELACIÓN PARA LOS RESIDUOS DEL CECGNVA



Gráfica 39. FUNCIÓN DE AUTOCORRELACIÓN PARA LOS RESIDUOS DEL CECR

En las graficas 38 y 39 de autocorrelacion de residuos para el consumo de energía muestra que cada una de las variables presentan aleatoriedad, por lo tanto los modelos escogidos para las distintas variables son los adecuados.

3.8 UTILIZACIÓN DEL MODELO PARA GENERAR PRONÓSTICOS

3.8.1 UTILIZACIÓN DEL MODELO PARA DATOS DEL NÚMERO DE ABONADOS

Tabla 12. PRONÓSTICOS DEL NÚMERO DE ABONADO PARA EL AÑO 2013

Mes	General		Residencial	Mes	General		Residencial
	Baja	Media	Baja		Baja	Media	Baja
Ene	19166	909	137312	Jul	19452	970	139545
Feb	19214	919	137684	Ago	19500	980	139917
Mar	19261	930	138056	Sep	19547	990	140289
Abr	19309	940	138428	Oct	19595	1000	140661
May	19357	950	138800	Nov	19643	1011	141033
Jun	19404	960	139172	Dic	19690	1021	141405

Elaborado por: Valeria Vacacela

Tabla 13. MEDIDAS DESCRIPTIVAS DE LOS PRONÓSTICOS DEL NÚMERO DE ABONADOS PARA EL AÑO 2013

Estadísticos	General		Residencial
	Baja	Media	
Media	19428,1	965,03	139358,53
Mediana	19428,1	965,03	139358,53
Desviación estándar	171,94	36,51	1341,63
Varianza de la muestra	29561,7	1333,1	1799979,29
Mínimo	19165,8	909,33	137311,97
Máximo	19690,4	1020,7	141405,09

Elaborado por: Valeria Vacacela

Los pronósticos generados indican que para el año 2013 se tiene en promedio 19428,14 abonados para la categoría general en el nivel de voltaje bajo y 965,03 abonados para la categoría general en el nivel de voltaje medio, un promedio de 139358,53 abonados para la categoría residencial.

3.8.2 UTILIZACIÓN DEL MODELO PARA DATOS DEL CONSUMO DE ENERGÍA EN MWh

Tabla 14. PRONOSTICO DEL CONSUMO DE ENERGÍA EN MWH PARA EL AÑO 2013

Mes	Alumbrado Público	General			Residencial
		Baja	Media	Alta	
Ene	2315,64	3678,08	5185,05	2782,62	9180,17
Feb	2313,01	3716,31	5237,21	2782,62	9217,76
Mar	2309,95	3755,34	5289,9	2782,62	9255,34
Abr	2306,46	3795,17	5343,12	2782,62	9292,93
May	2302,53	3835,8	5396,87	2782,62	9330,52
Jun	2298,18	3877,24	5451,14	2782,62	9368,1
Jul	2293,4	3919,48	5505,95	2782,62	9405,69
Ago	2288,19	3962,51	5561,29	2782,62	9443,28
Sep	2282,54	4006,35	5617,16	2782,62	9480,86
Oct	2276,47	4050,99	5673,56	2782,62	9518,45
Nov	2269,96	4096,44	5730,49	2782,62	9556,04
Dic	2263,03	4142,69	5787,95	2782,62	9593,62

Elaborado por: Valeria Vacacela

Tabla 15. MEDIDAS DESCRIPTIVAS DE LOS PRONÓSTICOS DEL CONSUMO DE ENERGÍA EN MWh PARA EL AÑO 2013.

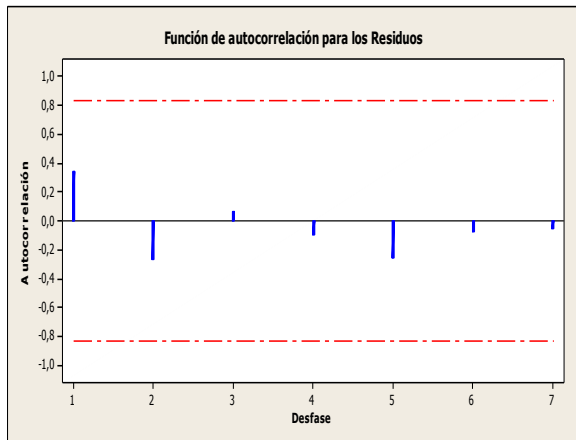
Estadísticos	Alumbrado Público	General			Residencial
		Baja	Media	Alta	
Media	2293,28	3903,03	5481,64	2782,62	9386,9
Error típico	5,03	43,98	57,05	0	39,12
Mediana	2295,79	3898,36	5478,55	2782,62	9386,9
Desviación estándar	17,41	152,35	197,64	0	135,52
Varianza de la muestra	303,01	23210,72	39060,55	0	18365,76
Mínimo	2263,03	3678,08	5185,05	2782,62	9180,17
Máximo	2315,64	4142,69	5787,95	2782,62	9593,62

Elaborado por: Valeria Vacacela

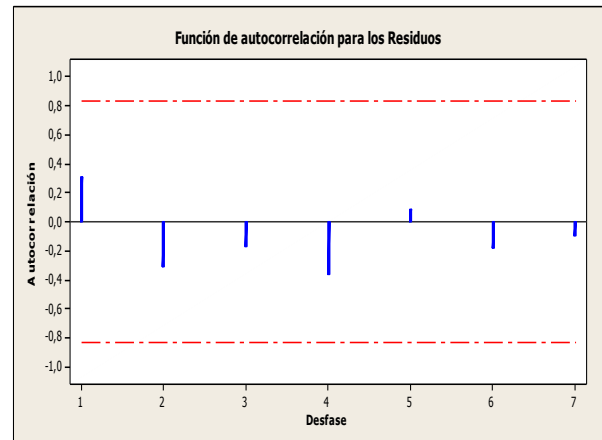
Se estimó que para el año 2013 el consumo medio fue de 2293,28 MWh en alumbrado público, 3903,03MWh en la categoría general nivel de voltaje bajo, 5481,64 MWh en la categoría general nivel de voltaje medio y 2782,62 MWh en la categoría general nivel de voltaje alto, para la categoría residencial se estima que el consumo medio fue de 9386,9 MWh.

3.9 VALIDACIÓN DEL MODELO

3.9.1 VALIDACIÓN DEL MODELO PARA DATOS DEL NÚMERO DE ABONADOS.

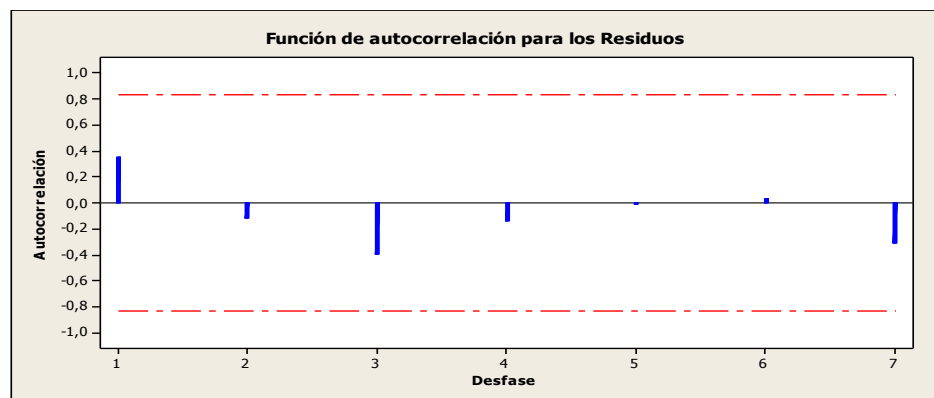


Gráfica 40. RESIDUOS DEL NACGNVB



Gráfica 41. RESIDUOS DEL NACGNVM

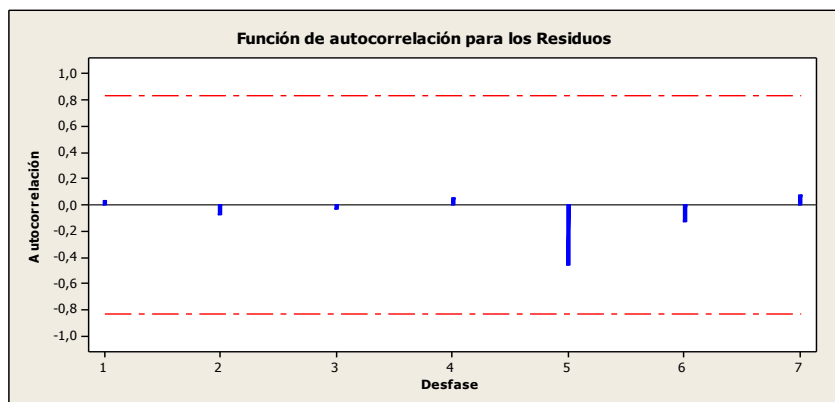
Las gráfica 40 y 41 indica que existe una aleatoriedad en los residuos de los valores reales con sus respectivos pronósticos, lo que corrobora la utilización del modelo de Holt para variable número de abonados categoría general nivel de voltaje bajo y medio.



Gráfica 42: RESIDUOS DEL NACR

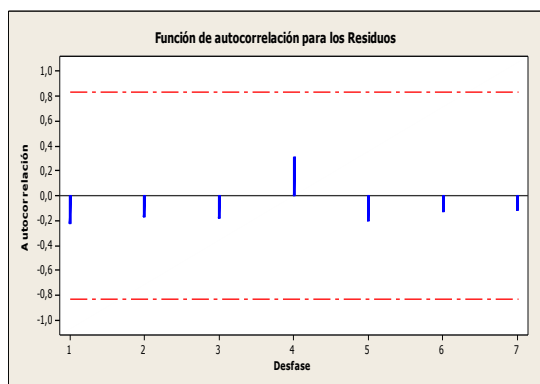
La aleatoriedad de los residuos que muestran en la gráfica 42 nos ayuda a corroborar que el modelo de pronóstico de Holt es el adecuado.

3.9.2 VALIDACIÓN DEL MODELO PARA DATOS DEL CONSUMO DE ENERGÍA EN MWh

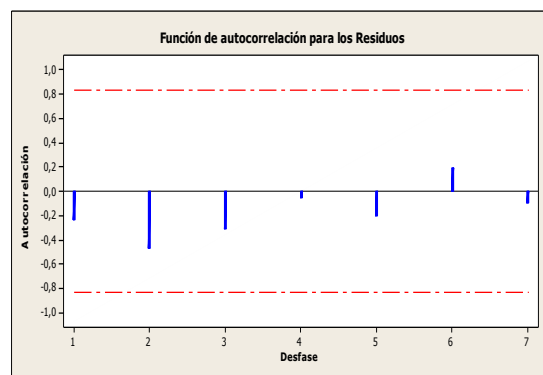


Gráfica 43: RESIDUOS VARIABLE CECAP

La gráfica 43 indica los residuos de los pronósticos con los valores reales del 2013 en donde existe aleatoriedad, por lo tanto el modelo de pronóstico cuadrático es el adecuado para la variable alumbrado público.



Gráfica 44: RESIDUOS DEL CECGNVB



Gráfica 45: RESIDUOS DEL CECGNVM

Los residuos del consumo de energía de las graficas 44 y 45 son aleatorios lo que indica que los pronósticos de las variables categoría general en el nivel de voltaje bajo y medio son adecuados con el modelo cuadrático.

Los autocorrelogramas que se han realizado para cada variable indican que existen aleatoriedad, por lo tanto el modelo de pronóstico Brown para el consumo de energía en la categoría general nivel de voltaje alto y el modelo de pronóstico de Holt para la variable residencial son los adecuados.

3.10 PRUEBA DE HIPÓTESIS

Hipótesis de Investigación

Los modelos elaborados para ser pronósticos de cada categoría del número de abonados y consumo de energía son significativos.

3.10.1 PRUEBA DE HIPÓTESIS PARA EL NÚMERO DE ABONADOS DE LA CATEGORÍA GENERAL NIVEL DE VOLTAJE MEDIO

PASO 1. Formulación de hipótesis

$$H_0: \mu_{reales} = \mu_{pronóstico}$$

$$H_1: \mu_{reales} \neq \mu_{pronóstico}$$

H_0 : los datos reales son estadísticamente iguales a los datos pronosticados mediante el modelo de Holt.

H_1 : los datos reales no son estadísticamente iguales a los datos pronosticados mediante el modelo de Holt.

PASO 2. Definición del nivel de significancia

La investigación se realizara con un $\alpha= 0.05$.

PASO 3. Tamaño de la muestra y cálculos estadísticos

Se tomaron los datos de enero a octubre, es decir $n= 8$ para los datos reales y $n=8$ para los pronósticos. Puesto que se desconoce la desviación estándar poblacional de los datos reales y de los pronósticos entonces se aplicara el estadístico t de student.

Tabla 16. NACGNVM REALES Y PRONÓSTICOS

Mes	General Medio		Mes	General Medio	
	Reales	Pronósticos		Reales	Pronósticos
Ene	895	909	May	941	950
Feb	903	919	Jun	945	960
Mar	928	930	Jul	951	970
Abr	962	940	Ago	952	980

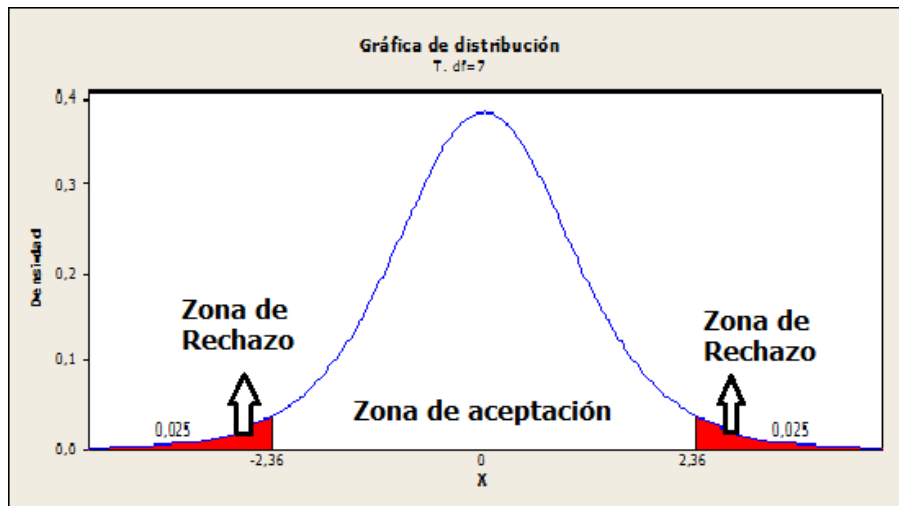
Elaborado por: Valeria Vacacela

Tabla 17. PRUEBA T PARA EL NACGNVM

	<i>Reales</i>	<i>Pronósticos</i>
Media	934,63	944,77
Varianza	583,13	615,31
Observaciones	8	8
Grados de libertad	7	
Estadístico t	-1,89	
P(T<=t) una cola	0,05	
Valor crítico de t (una cola)	1,89	
P(T<=t) dos colas	0,1	
Valor crítico de t (dos colas)	2,36	

Elaborado por: Valeria Vacacela

PASO 4. Definición de regiones de aceptación y rechazo.



PASO 5. Decisión estadística

Tomando en cuenta la gráfica anterior se tiene que el estadístico $t = -1,89$ está dentro de la zona de aceptación, es decir, es mayor que el valor crítico $-2,36$, entonces no se rechaza H_0 lo que significa que los valores reales son estadísticamente iguales a los pronosticados.

3.10.2 PRUEBA DE HIPÓTESIS PARA EL NÚMERO DE ABONADOS DE LA CATEGORÍA RESIDENCIAL.

PASO 1. Formulación de hipótesis

$$H_0: \mu_{reales} = \mu_{pronostico}$$

$$H_1: \mu_{reales} \neq \mu_{pronostico}$$

H_0 : los datos reales son estadísticamente iguales a los datos pronosticados mediante el modelo de Holt.

H_1 : los datos reales no son estadísticamente iguales a los datos pronosticados mediante el modelo de Holt.

PASO 2. Definición del nivel de significancia

Para nuestra investigación se toma un $\alpha=0,05$.

PASO 3. Tamaño de la muestra y cálculos estadísticos

Se tiene $n= 8$ en los datos reales y un $n=8$ en los pronósticos. Puesto que se desconoce la desviación estándar poblacional de los datos reales y pronósticos entonces se aplicara el estadístico t.

Tabla 18. NABCR REALES Y PRONÓSTICOS

Mes	Residencial		Mes	Residencial	
	Reales	Pronósticos		Reales	Pronósticos
Ene	137304	137312	May	138821	138800
Feb	137645	137684	Jun	139146	139172
Mar	138116	138056	Jul	139477	139545
Abr	138452	138428	Ago	139847	139917

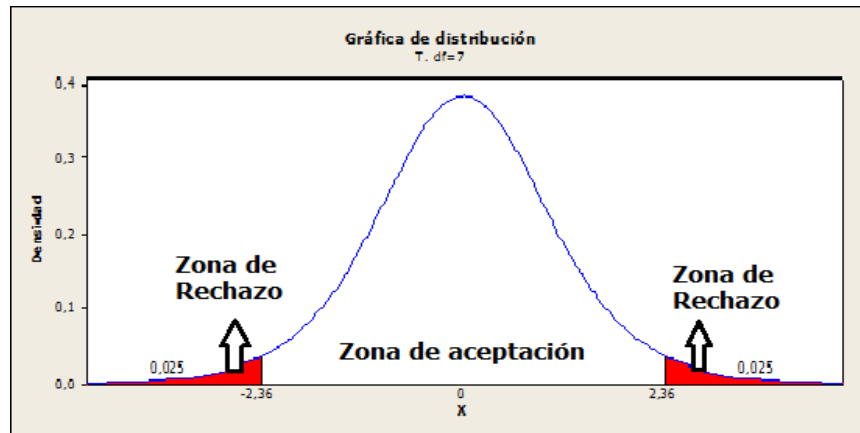
Elaborado por: Valeria Vacacela

Tabla 19. PRUEBA T PARA EL NACR

	<i>Reales</i>	<i>Pronósticos</i>
Media	138601	138614,33
Varianza	788412,57	830758,43
Observaciones	8	8
Grados de libertad	7	
Estadístico t	-0,82	
P(T<=t) una cola	0,22	
Valor crítico de t (una cola)	1,89	
P(T<=t) dos colas	0,44	
Valor crítico de t (dos colas)	2,36	

Elaborado por: Valeria Vacacela

PASO 4. Definición de regiones de aceptación y rechazo



PASO 5. Decisión estadística.

El estadístico $t = -0,82$ es mayor que el valor crítico $-2,36$, está dentro de la zona de aceptación, es decir los valores reales son estadísticamente iguales de los pronosticados.

3.10.3 PRUEBA DE HIPÓTESIS PARA CONSUMO DE ENERGÍA EN MWh DE LA CATEGORÍA GENERAL NIVEL DE VOLTAJE MEDIO

PASO 1. Formulación de hipótesis

$$H_0: \mu_{reales} = \mu_{pronóstico}$$

$$H_1: \mu_{reales} \neq \mu_{pronóstico}$$

H_0 : los datos reales son estadísticamente iguales a los datos pronosticados mediante el modelo cuadrático.

H_1 : los datos reales no son estadísticamente iguales a los datos pronosticados mediante el modelo de cuadrático.

PASO 2. Definición del Nivel de significancia

Para la prueba t se establece un $\alpha=0,05$.

PASO 3. Tamaño de la muestra y cálculos estadísticos

El tamaño de la muestra para los datos reales es de $n=8$ y para los pronosticados es de $n=8$. Puesto que se desconoce la desviación estándar poblacional de reales y pronósticos entonces se aplicara el estadístico t.

Tabla 20. CECGNVM REALES Y PRONÓSTICOS

Mes	General Medio		Mes	General Medio	
	Reales	Pronósticos		Reales	Pronósticos
Ene	5399,44	5185,05	May	5444,45	5396,87
Feb	4682,49	5237,21	Jun	5030,68	5451,14
Mar	4981,43	5289,9	Jul	5312,54	5505,95
Abr	5428,83	5343,12	Ago	5206,56	5561,29

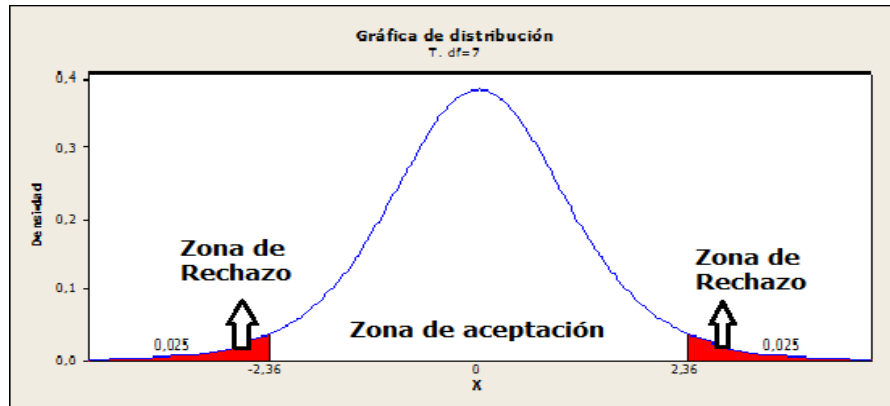
Elaborado por: Valeria Vacacela

Tabla 21. PRUEBA T PARA EL CECGNVM REALES Y PRONÓSTICOS

	<i>Reales</i>	<i>Pronósticos</i>
Media	5185,8	5371,32
Varianza	72464,25	17334,9
Observaciones	8	8
Grados de libertad	7	
Estadístico t	-1,92	
P(T<=t) una cola	0,05	
Valor crítico de t (una cola)	1,89	
P(T<=t) dos colas	0,1	
Valor crítico de t (dos colas)	2,36	

Elaborado por: Valeria Vacacela

PASO 4. Definición de regiones de aceptación y rechazo



PASO 5. Decisión estadística

El estadístico de prueba cae dentro de la zona de aceptación, ya que $-1,92 > -2,36$ entonces no se rechaza H_0 lo que significa que los datos pronosticados y los datos reales son estadísticamente iguales.

3.10.4 PLANTEAMIENTO DE HIPÓTESIS PARA CONSUMO DE ENERGÍA EN MWh DE LA CATEGORÍA GENERAL NIVEL DE VOLTAJE ALTO

PASO 1. Formulación de hipótesis

$$H_0: \mu_{reales} = \mu_{pronóstico}$$

$$H_1: \mu_{reales} \neq \mu_{pronóstico}$$

H_0 : los datos reales son estadísticamente iguales a los datos pronosticados mediante el modelo de Brown.

H_1 : los datos reales no son estadísticamente iguales a los datos pronosticados mediante el modelo de Brown.

PASO 2. Definición del nivel de significancia

Para la investigación se utiliza un $\alpha=0,05$.

PASO 3. Tamaño de la muestra y cálculos estadísticos

Para los datos reales se tiene una muestra $n= 8$ y para los datos de pronósticos $n=8$.Puesto que se desconoce la desviación estándar poblacional de los datos reales y pronósticos se aplicara el estadístico t.

Tabla 22. CECGNVA REALES Y PRONÓSTICOS

Mes	General Alto		Mes	General Alto	
	Reales	Pronósticos		Reales	Pronósticos
Ene	2785,55	2782,62	May	3709,92	2782,62
Feb	2365,32	2782,62	Jun	3442,27	2782,62
Mar	3050,75	2782,62	Jul	4015,17	2782,62
Abr	2820,25	2782,62	Ago	4307,71	2782,62

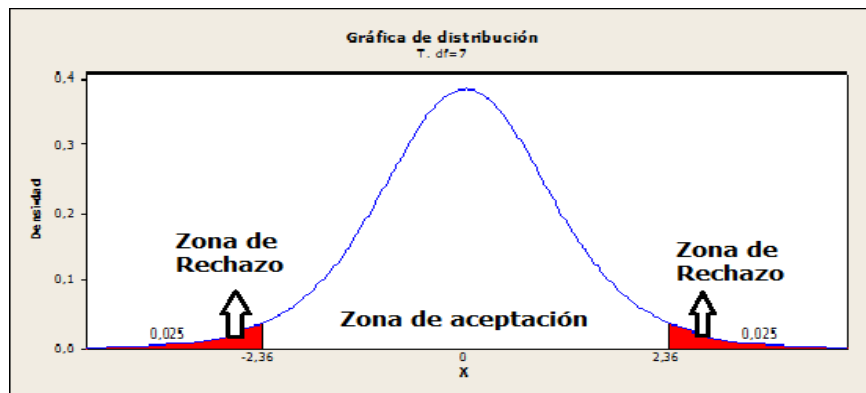
Elaborado por: Valeria Vacacela

Tabla 23.: PRUEBA T PARA EL CECGNVA REALES Y PRONÓSTICOS

	Reales	Pronósticos
Media	3312,12	2782,62
Varianza	449231,31	8,91E-07
Observaciones	8	8
Grados de libertad	7	
Estadístico t	2,23	
P(T<=t) una cola	0,03	
Valor crítico de t (una cola)	1,89	
P(T<=t) dos colas	0,06	
Valor crítico de t (dos colas)	2,36	

Elaborado por: Valeria Vacacela.

PASO 4. Definición de regiones de aceptación y rechazo



PASO 5. Decisión estadística

El estadístico $t = 2,23$ es menor que el valor crítico 2,36, por lo tanto el estadístico está dentro de la zona de aceptación, no se rechaza H_0 , es decir, los pronósticos y los datos reales son estadísticamente iguales.

CAPÍTULO IV

CONCLUSIONES Y RECOMENDACIONES

4.1 CONCLUSIONES.

- El AED identificó la presencia de dos datos atípicos en la variable consumo de energía MWh categoría general en el nivel de voltaje alto, los mismos que fueron depurados.
- La variable consumo de energía MWh categoría general nivel de voltaje bajo no siguen de una distribución normal y fue transformada.
- El consumo de energía MWh categoría general en el nivel voltaje alto presenta estacionariedad y las demás variables presentan tendencia y ciclicidad.
- Los modelos fueron identificados como los mejores por las medidas de precisión dando como resultados un modelo de Holt para las variables número de abonados categoría general en el nivel de voltaje bajo con el 0,24% de error, número de abonados categoría general en el nivel de voltaje medio con el 3,93% de error, número de abonados categoría residencial con el 0,1% de error y consumo de energía MWh categoría residencial con el 3% de error ; un modelo cuadrático para las variables consumo de energía MWh categoría alumbrado con el 3,46% de error , consumo de energía MWh, categoría general en el nivel de voltaje bajo con el 2,7% de error y consumo de energía MWh categoría general en el nivel de voltaje medio con el 4% de error; y un modelo de Brown para la variable consumo de energía MWh categoría general en el nivel de voltaje alto con el 10% de error.

- Los autocorrelogramas permitieron comprobar que los modelos de pronósticos son los adecuados.

4.2 RECOMENDACIONES.

- Con la finalidad de abastecer el servicio eléctrico que brinda la EERSA, se debería utilizar los modelos de pronósticos propuestos en el presente trabajo.
- Mejorar la recopilación de datos para facilitar los análisis estadísticos y econométricos.
- Actualizar los modelos de pronósticos periódicamente para una mejor toma de decisiones.

RESUMEN

Se realizó la modelación y predicción del **número de abonados** y del **consumo de energía en Megawatt - hora (MWh)** en función de **categorías** y **nivel de voltaje**, elaborado mediante la aportación de la base de datos del departamento de Planificación de la EERSA, con la finalidad de la obtención de modelos de pronósticos significativos para el conocimiento futuro del consumo de energía y el número de abonados.

Se partió de una base de datos mensuales, con cinco variables de **consumo de energía** y tres variables del **número de abonados** con 48 datos respectivamente, recurriendo al Análisis Exploratorio de Datos de dicha información se identificó la presencia de datos atípicos que fueron depurados, se aplicó la prueba de normalidad obteniendo una variable que no provenía de una distribución normal, siendo transformada por logaritmo natural. Las gráficas de series temporales y autocorrelogramas identificó la presencia de estacionariedad, tendencia y ciclicidad; aplicando las medidas de precisión MAPE, MAD y MSE, se obtuvo el mejor modelo de pronóstico.

Como resultado se obtuvo: un modelo de Holt para las variables número de abonados categoría general en el nivel de voltaje bajo, número de abonados categoría general en el nivel de voltaje medio, número de abonados categoría residencial y consumo de energía MWh categoría residencial; un modelo cuadrático para las variables consumo de energía MWh categoría alumbrado, consumo de energía MWh, categoría general en el nivel de voltaje bajo y consumo de energía MWh categoría general en el nivel de voltaje medio; y un modelo de Brown para la variable consumo de energía MWh categoría general en el nivel de voltaje alto. Los modelos de pronóstico que se establecieron fueron validados con los datos reales del 2013 por medio de la función de autocorrelación de residuos, indicando aleatoriedad el cual verifica que los modelos de pronósticos son los adecuados.

La EERSA debería instituir un formato apropiado para el almacenamiento de datos, de tal manera que los registros de consumo y el número de abonados se presenten de forma ordenada y comprensible, para facilitar un futuro análisis estadístico y econométrico.

SUMMARY

Modeling and prediction of the **number of customers, energy consumption** was carried out in **Megawatt - hour (MWh)** depending on **voltage level categories** and prepared by providing the database of the Planning Department in Empresa Eléctrica Riobamba S.A (EERSA), with the aim of obtaining forecast significant models for future knowledge of energy consumption and number of customers.

It started with a monthly database with five variables of energy consumption and three variables of the number of customers with 48 data respectively, using the Exploratory Data Analysis of the information, it was identified the presence of outliers that were released, normality test was applied to obtain a variable that did not come from a normal distribution, being transformed by natural logarithm. Graphs of time series and autocorrelations identified the presence of stationarity, trend and cyclicity using the measurements of MAPE, MAD and MSE position obtaining the best prediction model.

The obtained result was: a Holt model for the variables, number of customers in general category in low voltage level, number of customers in general category in medium voltage level, number of customers in residential category and energy consumption, MWh residential category; one quadratic model for energy consumption variables category MWh, lighting category in MWh energy consumption, general category at low voltage level and MWh energy, general category at medium voltage level; and a Brown model for the variable of MWh energy consumption in the general category of high voltage level.

Forecasting models established were varied with the actual data in 2013 by the autocorrelation function of waste, indicating randomness which verifies that forecast models are right.

The EERSA should institute an appropriate format for data storage, so that consumption records and the number of customers are presented in an orderly and understandable way to facilitate a future statistical and economic analysis.

BIBLIOGRAFÍA

BRIDGWATER Allan y Gill. Energías a Alternativas. 1ra ed., Madrid-España. Parqrinfo. 2009, pp. 120-125.

CONELEC. Estadísticas Sector Electrico Ecuatoriano., Quito-Ecuador. s. ed. 2012, pp 23-24.

<http://dspace.ups.edu.ec/bitstream/123456789/1882/3/02%20RECOPILACION%20DE%20INFORMACION.pdf>

2013/05/16

CONELEC. Pliego Tarifario para Empresas Eléctricas., Quito-Ecuador. s. ed. 2012, pp 5-7.

http://www.conelec.gob.ec/images/documentos/doc_10093_PLIEGO%20TARIFARIO%202012.pdf

2013/05/17

CONGACHA Jorge. Estadística Aplicada a la Educación con Actividades de Aprendizaje., Riobamba-Ecuador. s. ed. 2012, pp. 75-80, 244-260.

DETENCIÓN DE DATOS ATÍPICOS POR MEDIO DE LOS CUARTILRES. Ruiz J. 20013

<http://elestadistico.blogspot.com/2013/02/datos-atipicos-o-outliers.html>

2013/10/08

ENERGIA ELÉCTRICA EN EL ECUADOR. Sanchez Pedro. 20012

<http://es.slideshare.net/PedroSanchez14/generacion-transmicion-y-distribucion-de-energia-ecuador>

2013/05/15

FIGUERAS M., Salvador. Análisis Exploratorio de Datos., s. ed. 2003, pp. 10-40.

<http://ciberconta.unizar.es/Leccion/aed/ead.pdf>

2013/08/21

GALINDO., E. Estadística Métodos y Aplicaciones. 2. ed., Quito-Ecuador. s.ed. 2006, pp. 1-10, 33-37.

HANKE., John. Pronósticos en los Negocios. 9. ed, MéxicoD.F.-México. Pearson Educación. 2008, pp. 57-83, 101-133, 382.

Harper., E. Tecnologías de Generación de Energía Eléctrica., México D.F-México. Lumisa. 2011, pp. 16-17.

JOHNSON Richard. Probabilidad y Estadística para Ingenieros. 8. ed., México D.F-México. Pearson Educación. 2012, pp. 458-460.

MONTGOMERY., D. Probabilidad y Estadística Aplicada a la Ingeniería., México DF-México. s. ed. 1996, pp. 335-337, 363-364.

Murray R., Spiegel. Estadística. 4. ed., México D.F - México. Pearson Educación. 2009, pp. 246-286.

PÉREZ Cesar. Estadística., Madrid – España. Pearson Educación. 2003, pp. 381-398.

WALPOLE., M. Probabilidad y Estadística para Ingenieros. 8va ed., México D.F-México. s. ed. 2007, pp. 425-426.

ANEXOS

Anexo 1 MATRIZ DE DATOS DEL NÚMERO DE ABONADOS DEL PERIODO 2009-2012

Año	Mes	General		Residencial	Año	Mes	General		Residencial
		Baja	Media				Baja	Media	
2009	Ene	16757	472	116953	2011	Ene	17763	665	129199
	Feb	16783	477	117646		Feb	17764	723	129567
	Mar	16812	478	118080		Mar	17854	688	130081
	Abr	16875	483	118304		Abr	17953	687	130315
	May	17005	475	119075		May	17998	677	130586
	Jun	17028	480	119407		Jun	18117	706	130864
	Jul	16927	492	119429		Jul	18288	737	131066
	Ago	16963	488	119573		Ago	18336	742	131472
	Sep	17035	491	120307		Sep	18352	740	131620
	Oct	17113	490	121004		Oct	18436	743	131976
	Nov	17210	473	121786		Nov	18453	782	132385
	Dic	17290	458	122668		Dic	18505	780	132743
2010	Ene	17361	487	123506	2012	Ene	18693	786	133080
	Feb	17363	482	123608		Feb	18727	791	133519
	Mar	17389	482	123977		Mar	18755	792	133806
	Abr	17400	530	124469		Abr	18748	826	134176
	May	17428	507	124850		May	18807	840	134579
	Jun	17432	536	125476		Jun	18874	842	134874
	Jul	17552	536	126014		Jul	18910	849	135421
	Ago	17593	538	126672		Ago	19002	861	135796
	Sep	17684	860	127232		Sep	19023	859	136101
	Oct	17753	547	127960		Oct	19051	880	136362
	Nov	17599	718	128399		Nov	19088	896	136672
	Dic	17702	678	128733		Dic	19120	896	136963

Fuente: Ing. Enma Salazar

Anexo 2 MATRIZ DE DATOS DEL CONSUMO DE ENERGÍA EN MWH DEL PERIODO 2009-2012

Año	Mes	Alumbrado	General			Residencial	Año	Mes	Alumbrado	General			Residencial
		Público	Baja	Media	Alta				Público	Baja	Media	Alta	
2009	Ene	2050,53	2811,42	3294,93	2411	7640,98	2011	Ene	2426,81	3161,36	4447,31	2864	8978,97
	Feb	1956	2711,87	3229,04	2271	7397,12		Feb	2090,56	2783,26	3699,84	2695	7808,19
	Mar	1984,02	2714,53	3546,53	2249	7352,52		Mar	2272,62	2882,77	4204,06	2894	8453,23
	Abr	2117,19	2930,69	3326,02	1869	7918,44		Abr	2319,82	3077,39	3733,41	2627	8614,5
	May	2065,6	2821,13	3817,58	2817	7538,06		May	2310,52	3077,65	3877,02	2860	8567,42
	Jun	2074,17	2922,37	3540,99	2663	7620,46		Jun	2363,87	3113,84	4226,9	2845	8634,25
	Jul	2031,06	2796,46	3601,27	2733	7504,57		Jul	2265,27	3169,04	4232,62	2990	8263,74
	Ago	2058,06	2862,2	3370,67	2942	7594,13		Ago	2268,2	3032,14	4427,48	2798	8212,11
	Sep	2025,1	2815,05	3263,71	2773	7456,89		Sep	2397,1	3291,96	4404,27	2706	8654,94
	Oct	2104,74	2860,85	3612,19	2652	7752,07		Oct	2439,35	3233,61	4475,22	2319	8615,65
	Nov	1930,09	2655,64	3328,73	1329	7307,96		Nov	2368,3	3288,39	4533,76	2842	8503,96
	Dic	1978,93	2826,18	3250,28	2769	7433,87		Dic	2379,58	3344,61	4685,04	3100	8346,06
2010	Ene	2034,77	2692,98	3609,28	2772	7647,54	2012	Ene	2545,8	3444,71	4846,47	2864	8979,62
	Feb	2063	2871,24	3428,19	2588	7823,15		Feb	2428,75	3383,63	4655,09	2777,62	8711,55
	Mar	2005,26	2707,7	3870,45	2692	7432,28		Mar	2435,05	3349,77	4908,29	3015,72	8699,24
	Abr	2196,1	2953,52	3946,95	2363	8119,73		Abr	2215,75	3402,73	4734,44	2794	8980,53
	May	2148,36	2954,1	3912,64	1883,33	8017,03		May	2201,55	3363,78	4839,56	3096,02	8788,08
	Jun	2218,95	2993,75	3793,54	3014,78	8153,14		Jun	2207,86	3721,45	4926,7	2903,83	8757,02
	Jul	2166,61	2893,67	3869,19	3071,67	7912,8		Jul	2356,25	3521,72	5097,45	2837,86	9467,41
	Ago	2069,35	2796,49	3889,13	2917,33	7490,75		Ago	2277,29	3451,8	4971,11	2779,37	8930,55
	Sep	2207,56	2963,45	4320,89	2861,56	8009,64		Sep	2238,4	3428,91	4865,16	2916,64	8921,3
	Oct	2208,65	2981,98	3889,59	3098,28	8154,21		Oct	2292,71	3489,47	5213,61	2585,54	9037,7
	Nov	2263,72	2866,6	4307,39	2391,92	8361,14		Nov	2341,48	3659,48	4793,32	2614,83	9455,56
	Dic	2194,62	2849,14	4231,5	2927,92	8028,11		Dic	2216,7	3386,17	4888,37	2688,92	8934,98

Fuente: Ing. Enma Salazar

Anexo 3 MATRIZ DE DATOS DEL NÚMERO DE ABONADOS DEL AÑO 2013

Año	Mes	General		Residencial
		Baja	Media	
2013	Ene	19164	895	137304
	Feb	19215	903	137645
	Mar	19241	928	138116
	Abr	19231	962	138452
	May	19316	941	138821
	Jun	19362	945	139146
	Jul	19388	951	139477
	Ago	19452	952	139847
	Sep	20208	201	140215
	Oct	20256	201	140421
	Nov	20350	200	140659
	Dic	20383	200	140859

Anexo 4 MATRIZ DE DATOS DEL CONSUMO DE ENERGÍA DEL AÑO 2013

Año	Mes	Alumbrado Público	General			Residencial
			Baja	Media	Alta	
2013	Ene	2567,4	3791,21	5399,44	2785,55	10065,68
	Feb	2279,39	3448,23	4682,49	2365,32	9355,4
	Mar	2386,82	3638,7	4981,43	3050,75	9458,31
	Abr	2439,81	3725,28	5428,83	2820,25	9547,48
	May	2518,94	3842,36	5444,45	3709,92	9788,03
	Jun	2379,85	3505,71	5030,68	3442,27	9416,91
	Jul	2589,09	3706,12	5312,54	4015,17	10623,65
	Ago	2581,79	3786,76	5206,56	4307,71	10067,7
	Sep	2350,82	4779,07	3736,39	2815,51	9524,07
	Oct	2513,67	4994,27	4248,54	3170,9	9838,2
	Nov	2702	5308,67	3867,07	3853,8	10643,65
	Dic	2518,61	4962,52	3804,67	4405,77	10144,25